O'Bryen, Barbara

From: Zhou, Shubo (AU1631)
Sent: Wednesday, December 26, 2001 9:55 AM
To: O'Bryen, Barbara
Subject: search request for 09/422,838

Happy holiday, Barb! Another search for you. Enjoy!

Joe

Shubo "Joe" Zhou, Ph.D.
Patent Examiner
(703)-605-1158, CM1/12B03
AU 1631, US PTO

Search Request

****

Requester's full name: *Shubo "Joe" Zhou* ***Examiner #:*** *78282*

Art Unit: *1631* ***Phone #:*** *703-605-1158* ***Mailbox #:*** *12D01/CM1*

Results format: *pape* ***Room #:*** *12B03*

****

Serial #: 09/422,838

Please search:

Protein databases for
SEQ ID NO: 33

Including:
1. default search

Please provide 45 alignments for the search.

POINT OF CONTACT:
BARB O'BRYEN
TECH. INFORMATION SPECIALIST
STIC CM1 12C14 308-4291

BOB
12-26-01

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:26:08 ; Search time 13.48 Seconds
(without alignments)
203.434 Million cell updates/sec

Title: US-09-422-838C-33
Perfect score: 197
Sequence: 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	67	34.0	865	2	T34584	probable secreted
2	63.5	32.2	209	2	B42687	neurotrophin-4 pre
3	60.5	30.7	210	2	A42687	neurotrophin-4 pre
4	60	30.5	500	2	T20961	hypothetical prote
5	59.5	30.2	518	2	S72938	hflX protein - Myc
6	59	29.9	683	2	B71325	conserved hypothet
7	58	29.4	302	2	S71334	acetyl xylan ester
8	58	29.4	924	2	E71476	alanine--tRNA liga
9	57.5	29.2	495	2	D70505	probable HflX - My
10	57	28.9	339	2	S20880	homeotic protein H
11	56.5	28.7	77	1	INSH	insulin precursor
12	56.5	28.7	105	1	IPBO	insulin precursor
13	56.5	28.7	176	2	B72698	hypothetical prote
14	56	28.4	56	2	T12783	sublancin 168 prec
15	56	28.4	163	2	T33130	hypothetical prote
16	56	28.4	349	2	E86405	hypothetical prote
17	56	28.4	510	2	S41943	cellulose 1,4-beta
18	56	28.4	511	2	S44716	cellulose 1,4-beta
19	56	28.4	540	2	S41942	cellulose 1,4-beta
20	56	28.4	619	1	KSNCLO	laccase (EC 1.10.3
21	56	28.4	619	1	KSNCLT	laccase (EC 1.10.3
22	56	28.4	767	2	E70895	hypothetical glyci
23	55	27.9	180	2	T49530	related to glycine
24	55	27.9	201	2	T49792	hypothetical prote
25	55	27.9	257	2	C84890	hypothetical prote
26	55	27.9	331	2	T26807	hypothetical prote
27	55	27.9	333	2	T26808	hypothetical prote
28	55	27.9	393	2	T20268	hypothetical prote
29	55	27.9	399	2	T47712	MYB transcription
30	55	27.9	556	2	A32466	numb protein - fru
31	55	27.9	875	2	H81739	alanyl-tRNA synthe
32	54.5	27.7	112	2	F70954	probable lsr2 prot
33	54.5	27.7	198	2	A54507	dnaK-type molecula
34	54.5	27.7	246	2	G02371	U1-snRNP binding p
35	54.5	27.7	1028	2	A56038	DNA-binding protei
36	54.5	27.7	1213	2	S16356	ovo protein - frui
37	54	27.4	201	2	JQ1094	hypothetical 20.2K
38	54	27.4	445	1	A49447	transcription fact
39	54	27.4	490	2	T09084	phosphatidylinosit
40	54	27.4	495	2	E70948	probable amidase -
41	54	27.4	620	2	F64408	coenzyme F420 hydr
42	54	27.4	1001	2	T13807	potassium channel
43	53.5	27.2	562	2	F72771	probable lysyl-tRN
44	53	26.9	261	2	T37948	probable U1 small
45	53	26.9	309	2	T19389	hypothetical prote
46	53	26.9	497	2	T45406	probable amidase [
47	53	26.9	3190	2	T13828	CREB-binding prote
48	52.5	26.6	65	2	T48968	glycine-rich prote
49	52.5	26.6	198	2	A57717	transcription fact
50	52.5	26.6	341	1	PYVZCB	spheroidin precurs
51	52.5	26.6	362	2	H75398	probable succinyl-
52	52.5	26.6	448	2	A36311	70K U1 small nucle
53	52.5	26.6	487	2	B39490	subtilisin-like pr
54	52.5	26.6	514	2	A35658	transcription fact
55	52.5	26.6	652	1	JC2191	subtilisin-like pr
56	52.5	26.6	786	2	A47546	triacylglycerol li
57	52.5	26.6	787	2	T05617	hypothetical prote
58	52.5	26.6	839	2	B96576	hypothetical prote
59	52.5	26.6	904	2	A84212	hypothetical prote
60	52.5	26.6	962	2	JC5571	subtilisin-like pr
61	52.5	26.6	969	1	A39490	subtilisin-like pr
62	52.5	26.6	975	2	JC5570	subtilisin-like pr
63	52	26.4	63	2	T31193	hypothetical prote
64	52	26.4	284	2	S74256	homeotic protein s
65	52	26.4	330	2	S74255	homeotic protein s
66	52	26.4	415	2	D96664	hypothetical prote
67	52	26.4	424	2	T01383	GTPase-activating
68	52	26.4	426	2	T04318	homeobox protein L
69	52	26.4	443	1	S29334	transcription fact
70	52	26.4	445	1	S31224	transcription fact
71	52	26.4	448	2	S15018	transcription fact
72	52	26.4	465	2	S41644	polyadenylate-bind
73	52	26.4	494	2	F70856	probable gatA - My
74	52	26.4	545	1	COBYC2	cyclin 2 - yeast (
75	52	26.4	546	1	COBYC1	cyclin 1 - yeast (
76	52	26.4	593	1	KRHU0	keratin 10, type I
77	52	26.4	640	2	T08179	LRG5 protein - Chl
78	52	26.4	777	2	S65543	3',5'-cyclic-nucle
79	52	26.4	1168	1	MWAXIC	myosin heavy chain
80	52	26.4	1176	2	A49848	nitrite reductase
81	51.5	26.1	371	2	T13021	hypothetical prote
82	51.5	26.1	378	2	S04336	U1 snRNP 70K prote
83	51.5	26.1	440	2	S71795	transcription fact
84	51.5	26.1	471	2	S02016	U1 snRNP 70K prote
85	51.5	26.1	614	2	A25707	U1 snRNP 70K prote
86	51.5	26.1	864	2	A48266	protein-tyrosine k
87	51	25.9	103	2	T47718	hypothetical prote
88	51	25.9	171	2	H84709	probable glycine-r
89	51	25.9	237	2	B82986	hypothetical prote
90	51	25.9	249	2	T04436	ankyrin 3 homolog
91	51	25.9	285	2	S69312	probable membrane
92	51	25.9	295	2	E84862	hypothetical prote
93	51	25.9	303	2	S71185	splicing factor SF
94	51	25.9	306	2	D70601	UTP--glucose-1-pho
95	51	25.9	323	2	S20099	transforming prote
96	51	25.9	385	2	T20410	hypothetical prote
97	51	25.9	475	2	A43915	homeotic protein e
98	51	25.9	482	2	D75346	glutamyl-tRNA(Gln)
99	51	25.9	495	2	T52066	probable farnesyl
100	51	25.9	497	2	T35116	hypothetical prote

ALIGNMENTS

```
RESULT   1
T34584
probable secreted proteinase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 24-Nov-1999
C;Accession: T34584
R;Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z21548
A;Accession: T34584
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-865 <MUR>
A;Cross-references: EMBL:AL021529; PIDN:CAA16449.1; GSPDB:GN00070; SCOEDB:SC10A5.17
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC10A5.17


  Query Match             34.0%;  Score 67;  DB 2;  Length 865;
  Best Local Similarity   66.7%;  Pred. No. 4.6;
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       9 WLAARAGGGCGGGGIEGP 26
           |||| | | |||||   |
Db     651 WLAACAAGNCGGGGTNPP 668


RESULT   2
B42687
neurotrophin-4 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
C;Accession: B42687; JH0504; JH0505
R;Ip, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio, L.
Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992
A;Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distribut
A;Reference number: A42687; MUID:92212967
A;Accession: B42687
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <IPA>
A;Cross-references: GB:M86742; NID:g205775; PIDN:AAA41728.1; PID:g205776
R;Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenthal,
Neuron 7, 857-866, 1991
A;Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.
A;Reference number: JH0503; MUID:92075279
A;Accession: JH0504
A;Molecule type: DNA
A;Residues: 1-209 <BER>
A;Accession: JH0505
A;Molecule type: mRNA
A;Residues: 1-176,'P',178-209 <BER1>
A;Cross-references: GB:S69323; NID:g240025; PIDN:AAB20548.1; PID:g240026
C;Comment: This protein is a targed-derived, diffusible neurotrophic factor.
C;Comment: The neurotrophins stimulate autophosphorylation and transduce signals through
C;Superfamily: nerve growth factor beta chain
C;Keywords: glycoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-79/Domain: propeptide #status predicted <PRO>
F;80-209/Product: neurotrophin-5 #status predicted <NEU>
F;75/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             32.2%;  Score 63.5;  DB 2;  Length 209;
  Best Local Similarity   37.5%;  Pred. No. 3.2;
  Matches   15;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    1;

Qy       3 GPTLRQWL---------AARAGGGCGGGGIEGPTLRQWLA 33
           |  |||:          |   | | ||||  |   | ||:
Db     128 GSPLRQYFFETRCKAESAGEGGPGVGGGGCRGVDRRHWLS 167


RESULT   3
A42687
neurotrophin-4 precursor - human
N;Alternate names: neurotrophin-5
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
C;Accession: A42687; JH0503
R;Ip, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio,
Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992
A;Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distri
A;Reference number: A42687; MUID:92212967
A;Accession: A42687
A;Molecule type: DNA
A;Residues: 1-210 <IP1>
A;Cross-references: GB:M86528; NID:g190264; PIDN:AAA60154.1; PID:g190265
A;Note: sequence extracted from NCBI backbone (NCBIN:93810, NCBIP:93811)
R;Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenth
Neuron 7, 857-866, 1991
A;Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.
A;Reference number: JH0503; MUID:92075279
A;Accession: JH0503
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-210 <BER>
C;Comment: The neurotrophins stimulate autophosphorylation and transduce signals thro
C;Genetics:
A;Gene: GDB:NTF5
A;Cross-references: GDB:134723; OMIM:162662
A;Map position: 19pter-19qter
C;Superfamily: nerve growth factor beta chain
C;Keywords: glycoprotein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-80/Domain: propeptide #status predicted <PRO>
F;81-210/Product: neurotrophin-4 #status predicted <NEU>
F;76/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             30.7%;  Score 60.5;  DB 2;  Length 210;
  Best Local Similarity   35.0%;  Pred. No. 6.9;
  Matches   14;  Conservative    3;  Mismatches   14;  Indels    9;  Gaps    1;

Qy       3 GPTLRQWL---------AARAGGGCGGGGIEGPTLRQWLA 33
           |  |||:          |   | | ||||  |   | |::
Db     129 GSPLRQYFFETRCKADNAEEGGPGAGGGGCRGVDRRHWVS 168


RESULT   4
T20961
hypothetical protein F15B9.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20961
R;Percy, C.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19351
A;Accession: T20961
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-500 <WIL>
A;Cross-references: EMBL:Z78013; PIDN:CAB01420.1; GSPDB:GN00023; CESP:F15B9.5
A;Experimental source: clone F15B9
C;Genetics:
A;Gene: CESP:F15B9.5
A;Map position: 5
A;Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1


  Query Match             30.5%;  Score 60;  DB 2;  Length 500;
  Best Local Similarity   52.2%;  Pred. No. 17;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;
```

```
Qy       3 GPTLRQWLAARAGGGCGGGGIEG 25
           |  | ::|: | ||| ||||: |
Db     429 GSMLGRFLSNRGGGGGGGGGMGG 451
```

RESULT 5
S72938
hflX protein - Mycobacterium leprae
N;Alternate names: B2235_C2_202 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C;Accession: S72938
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B2235.
A;Reference number: S72587
A;Accession: S72938
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-518 <SMI>
A;Cross-references: EMBL:U00019; NID:g467079; PIDN:AAA17274.1; PID:g467091
C;Genetics:
A;Start codon: GTG
C;Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

```
  Query Match             30.2%;  Score 59.5;  DB 2;  Length 518;
  Best Local Similarity   43.3%;  Pred. No. 19;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    7;  Gaps    1;

Qy       4 PTLRQW-------LAARAGGGCGGGGIEGP 26
           | || |       :  |||| || |: ||
Db     219 PRLRGWGESMSRQVGGRAGGSGGGVGLRGP 248
```

RESULT 6
B71325
conserved hypothetical protein TP0421 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C;Accession: B71325
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770
A;Accession: B71325
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-683 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65409.1; PID:g33227
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0421

```
  Query Match             29.9%;  Score 59;  DB 2;  Length 683;
  Best Local Similarity   43.8%;  Pred. No. 28;
  Matches   14;  Conservative    2;  Mismatches   12;  Indels    4;  Gaps    1;

Qy       4 PTLRQWLAARAGGGCGGGGIEGPTLRQWLAAR 35
           | : :||    |      ||||  | || |||
Db      74 PLILEWL----GNAYYRSGIEGAALHQWGAAR 101
```

RESULT 7
S71334
acetyl xylan esterase precursor - fungus (Trichoderma reesei)
C;Species: Trichoderma reesei
C;Date: 23-Jul-1997 #sequence_revision 01-Aug-1997 #text_change 17-Mar-1999
C;Accession: S71334
R;Margolles-Clark, E.; Tenkanen, M.; Soederlund, H.; Penttilae, M.
Eur. J. Biochem. 237, 553-560, 1996
A;Title: Acetyl xylan esterase from Trichoderma reesei contains an active-site serine
A;Reference number: S71334; MUID:96235218
A;Accession: S71334
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <MAR>
A;Cross-references: EMBL:Z69256; NID:g1431619; PID:e220701; PID:g1431620
C;Genetics:
A;Gene: axe1
C;Superfamily: fungal cellulose-binding domain homology
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-302/Product: acetyl xylan esterase #status predicted <MAT>
F;271-302/Domain: fungal cellulose-binding domain homology <FCB>

```
  Query Match             29.4%;  Score 58;  DB 2;  Length 302;
  Best Local Similarity   35.9%;  Pred. No. 18;
  Matches   14;  Conservative    1;  Mismatches    8;  Indels   16;  Gaps    2;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT----------LRQW 31
           |||   |      | ||| |  |||          : ||
Db     265 GPTQTHW------GQCGGQGWTGPTQCESGTTCQVISQW 297
```

RESULT 8
E71476
alanine--tRNA ligase (EC 6.1.1.7) - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: E71476
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809
A;Accession: E71476
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-924 <ARN>
A;Cross-references: GB:AE001346; GB:AE001273; NID:g3329203; PIDN:AAC68344.1; PID:g332
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: alaS
C;Superfamily: alanine--tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

```
  Query Match             29.4%;  Score 58;  DB 2;  Length 924;
  Best Local Similarity   30.6%;  Pred. No. 47;
  Matches   15;  Conservative    5;  Mismatches   15;  Indels   14;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIE-------------GPTLRQWLAAR 35
           ::  ||   | |  || |||  |                |||||:: :
Db     874 VQAHTLLAELLAPYGGRCGGKAISAQGSSAELPQIEFLNKTLRQWISTQ 922
```

RESULT 9
D70505
probable HflX - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Sep-2000
C;Accession: D70505
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: D70505
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

```
A;Residues: 1-495 <COL>
A;Cross-references: GB:Z98209; GB:AL123456; NID:g3261838; PIDN:CAB10901.1; PID:e332282;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: hflX
C;Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology


  Query Match            29.2%;  Score 57.5;  DB 2;  Length 495;
  Best Local Similarity  43.3%;  Pred. No. 31;
  Matches   13;  Conservative    1;  Mismatches    9;  Indels    7;  Gaps    1;

Qy       4 PTLRQW-------LAARAGGGCGGGGIEGP 26
           | || |         ||||  || |: ||
Db     199 PRLRGWGESMSRQAGGRAGGSGGGVGLRGP 228


RESULT  10
S20880
homeotic protein Hox 4.5 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 17-Nov-2000
C;Accession: S20880; S09569; S09398
R;Renucci, A.; Zappavigna, V.; Zakany, J.; Izpisua-Belmonte, J.C.; Buerki, K.; Duboule,
EMBO J. 11, 1459-1468, 1992
A;Title: Comparison of mouse and human HOX-4 complexes defines conserved sequences invol
A;Reference number: S20879; MUID:92224884
A;Accession: S20880
A;Molecule type: DNA
A;Residues: 1-339 <REN>
A;Cross-references: EMBL:X62669; NID:g51414; PIDN:CAA44542.1; PID:g51416
R;Duboule, D.; Dolle, P.
EMBO J. 8, 1497-1505, 1989
A;Title: The structural and functional organization of the murine HOX gene family resemb
A;Reference number: S09569; MUID:89356621
A;Accession: S09569
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 272-331 <DUB>
A;Cross-references: EMBL:X14714; NID:g51427; PIDN:CAB57813.1; PID:g6015583
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1989
R;Dolle, P.; Duboule, D.
EMBO J. 8, 1507-1515, 1989
A;Title: Two gene members of the murine HOX-5 complex show regional and cell-type specif
A;Reference number: S09398; MUID:89356622
A;Accession: S09398
A;Molecule type: DNA
A;Residues: 272-331 <DOL>
A;Cross-references: GB:X14714; GB:M21040; NID:g51427; PIDN:CAB57813.1; PID:g6015583
C;Genetics:
A;Gene: Hox-4.5
A;Introns: 260/1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;273-329/Domain: homeobox homology <HOX>


  Query Match            28.9%;  Score 57;  DB 2;  Length 339;
  Best Local Similarity  40.6%;  Pred. No. 25;
  Matches   13;  Conservative    2;  Mismatches    9;  Indels    8;  Gaps    1;

Qy       3 GPTLRQWL--------AARAGGGCGGGGIEGP 26
           |  :| |:         |  ||| ||||  ||
Db     101 GRYVRSWMEPLPGFPGGAGGGGGSGGGGGGGP 132


RESULT  11
INSH
insulin precursor - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C;Accession: S16430; S16431
R;Brown, H.; Sanger, F.; Kitai, R.
Biochem. J. 60, 556-565, 1955
A;Title: The structure of pig and sheep insulins.
A;Reference number: A90344
A;Accession: S16430
A;Molecule type: protein
A;Residues: 1-30;57-77 <BRO>
R;Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin
A;Reference number: A92111; MUID:72258016
A;Accession: S16431
A;Molecule type: protein
A;Residues: 31-56 <PET>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCH>
F;1-30,57-77/Product: insulin #status experimental <MAT>
F;31-56/Domain: connecting peptide #status experimental <CPEP>
F;57-77/Domain: insulin chain A #status experimental <ACH>
F;7-63,19-76,62-67/Disulfide bonds: #status predicted


  Query Match            28.7%;  Score 56.5;  DB 1;  Length 77;
  Best Local Similarity  50.0%;  Pred. No. 7.8;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGP 26
           :|||   |  |    || | ||:|||
Db      32 VEGP---QVGALELAGGPGAGGLEGP 54


RESULT  12
IPBO
insulin precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 22-Apr-1995 #text_change 16-Jul-1999
C;Accession: A40909; A92080; A92074; A91185; A90342; A90341; S48184; S48185; S46258;
R;D'Agostino, J.; Younes, M.A.; White, J.W.; Besch, P.K.; Field, J.B.; Frazier, M.L.
Mol. Endocrinol. 1, 327-331, 1987
A;Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic a
A;Reference number: A40909; MUID:88288209
A;Accession: A40909
A;Molecule type: mRNA
A;Residues: 1-105 <DAA>
A;Cross-references: GB:M54979; NID:g163578; PIDN:AAA30722.1; PID:g163579
A;Experimental source: fetal pancreas
R;Nolan, C.; Margoliash, E.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 2780-2795, 1971
A;Title: The structure of bovine proinsulin.
A;Reference number: A92080; MUID:71166442
A;Accession: A92080
A;Molecule type: protein
A;Residues: 25-105 <NOL>
R;Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.
J. Biol. Chem. 246, 1365-1374, 1971
A;Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.
A;Reference number: A92074; MUID:71116409
A;Accession: A92074
A;Molecule type: protein
A;Residues: 57-82 <STE>
R;Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 183-189, 1971
A;Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancre
A;Reference number: A91185; MUID:71257721
A;Accession: A91185
A;Molecule type: protein
A;Residues: 57-82 <SAL>
R;Sanger, F.; Thompson, E.O.P.
Biochem. J. 53, 366-374, 1953
A;Title: The amino-acid sequence in the glycyl chain of insulin. 2. The investigation
A;Reference number: A90342
A;Accession: A90342
A;Molecule type: protein
```

```
A;Residues: 85-105 <SAN>
R;Sanger, F.; Tuppy, H.
Biochem. J. 49, 481-490, 1951
A;Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigat
A;Reference number: A90341
A;Accession: A90341
A;Molecule type: protein
A;Residues: 25-54 <SA2>
R;Cheng, R.; Kawakishi, S.
Eur. J. Biochem. 223, 759-764, 1994
A;Title: Site-specific oxidation of histidine residues in glycated insulin mediated by C
A;Reference number: S48184; MUID:94333378
A;Accession: S48184
A;Molecule type: protein
A;Residues: 85-105 <CHE>
A;Accession: S48185
A;Status: preliminary
A;Molecule type: protein
A;Residues: 25-30,'X',32-42,'X',44-54 <CH2>
R;Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.
Biochem. J. 60, 541-556, 1955
A;Title: The disulphide bonds of insulin.
A;Reference number: A90343
A;Contents: annotation; amides; disulfides
R;Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.
FEBS Lett. 349, 205-209, 1994
A;Title: Existence of a molecular ruler in proteasomes suggested by analysis of degradat
A;Reference number: S46258; MUID:94326921
A;Accession: S46258
A;Status: preliminary
A;Molecule type: protein
A;Residues: 25-54 <WEN>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54/Domain: insulin chain B #status experimental <BCH>
F;25-54,85-105/Product: insulin #status experimental <MAT>
F;57-82/Domain: connecting peptide #status experimental <CPEP>
F;85-105/Domain: insulin chain A #status experimental <ACH>
F;31-91,43-104,90-95/Disulfide bonds: #status experimental


  Query Match             28.7%;  Score 56.5;  DB 1;  Length 105;
  Best Local Similarity   50.0%;  Pred. No. 10;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGP 26
           :|||   |  |    || | ||:|||
Db      58 VEGP---QVGALELAGGPGAGGLEGP 80


RESULT  13
B72698
hypothetical protein APE1002 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: B72698
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: B72698
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-176 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79986.1; PID:d1043772; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1002


  Query Match             28.7%;  Score 56.5;  DB 2;  Length 176;
  Best Local Similarity   34.9%;  Pred. No. 16;
  Matches   15;  Conservative    1;  Mismatches    8;  Indels   19;  Gaps    1;

Qy       7 RQWLAARAGGGC------------------GGGGIEGPTLRQ 30
           || |    ||||                  |||| ||  ||:
Db      12 RQGLHGEEGGGCDCKGCGRRLNPPPPGHHWQGGGGEEGEELRR 54


RESULT  14
T12783
sublancin 168 precursor - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999
C;Accession: T12783; H69719
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 pro
A;Reference number: Z17583
A;Accession: T12783
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-56 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025497; PIDN:AAC12992.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: H69719
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-56 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14066.1; PID:e11835
A;Experimental source: strain 168
C;Genetics: <LA1>
A;Gene: yolG
C;Genetics: <KU1>
A;Gene: sunA
C;Superfamily: unassigned lanthionine-containing peptides
C;Keywords: antibiotic; lanthionine
F;1-19/Domain: propeptide #status predicted <PRO>
F;20-56/Product: sublancin 168 #status predicted <MAT>
F;26-55/Disulfide bonds: #status experimental
F;33-48/Disulfide bonds: #status predicted
F;35/Modified site: dehydroalanine (Ser) #status experimental
F;38-41/Cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted


  Query Match             28.4%;  Score 56;  DB 2;  Length 56;
  Best Local Similarity   44.0%;  Pred. No. 6.7;
  Matches   11;  Conservative    3;  Mismatches    9;  Indels    2;  Gaps    1;

Qy       9 WLAARAGG--GCGGGGIEGPTLRQW 31
           ||   :||  ||||| :     ||:
Db      30 WLQCASGGTIGCGGGAVACQNYRQF 54


RESULT  15
T33130
hypothetical protein C23H5.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
```

```
C;Accession: T33130
R;Lamar, E.; Kramer, J.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C23H5.
A;Reference number: Z21286
A;Accession: T33130
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-163 <LAM>
A;Cross-references: EMBL:AF067609; PIDN:AAC17537.1; GSPDB:GN00022; CESP:C23H5.9
A;Experimental source: strain Bristol N2; clone C23H5
C;Genetics:
A;Gene: CESP:C23H5.9
A;Map position: 4
A;Introns: 1/3; 101/3; 126/2


  Query Match             28.4%;  Score 56;  DB 2;  Length 163;
  Best Local Similarity   75.0%;  Pred. No. 17;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    1;

Qy      15 GGGCGGGGIEG--PTL 28
           ||||||||  |  |||
Db      33 GGGCGGGGGGGCLPTL 48


RESULT  16
E86405
hypothetical protein AAG26943.1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E86405
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: E86405
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-349 <STO>
A;Cross-references: GB:AE005172; NID:g11024859; PIDN:AAG26943.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1


  Query Match             28.4%;  Score 56;  DB 2;  Length 349;
  Best Local Similarity   66.7%;  Pred. No. 33;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      11 AARAGGGCGGGGIEG 25
           |:  ||||||||  |
Db     333 ASCGGGGCGGGGCGG 347


RESULT  17
S41943
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chrysospo
C;Species: Phanerochaete chrysosporium
C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S44715; S41943
R;Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
Mol. Microbiol. 12, 209-216, 1994
A;Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in the l
A;Reference number: S44714; MUID:94335641
A;Accession: S44715
A;Molecule type: mRNA
A;Residues: 1-510 <SI2>
A;Cross-references: EMBL:Z29653; NID:g453222; PIDN:CAA82762.1; PID:g453224
C;Genetics:
A;Introns: 505/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;479-510/Domain: fungal cellulose-binding domain homology <FCB>


  Query Match             28.4%;  Score 56;  DB 2;  Length 510;
  Best Local Similarity   48.0%;  Pred. No. 46;
  Matches   12;  Conservative    1;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT 27
           |||: ||      | ||| |  | |
Db     473 GPTVPQW------GQCGGIGYSGST 491


RESULT  18
S44716
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chryso
C;Species: Phanerochaete chrysosporium
C;Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C;Accession: S44716; S33165
R;Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
Mol. Microbiol. 12, 209-216, 1994
A;Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in th
A;Reference number: S44714; MUID:94335641
A;Accession: S44716
A;Molecule type: DNA
A;Residues: 1-511 <SIM>
A;Cross-references: EMBL:Z22527; NID:g296028; PIDN:CAA80252.1; PID:g3980202
C;Genetics:
A;Introns: 201/3; 506/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;480-511/Domain: fungal cellulose-binding domain homology <FCB>


  Query Match             28.4%;  Score 56;  DB 2;  Length 511;
  Best Local Similarity   48.0%;  Pred. No. 46;
  Matches   12;  Conservative    1;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT 27
           |||: ||      | ||| |  | |
Db     474 GPTVPQW------GQCGGIGYSGST 492


RESULT  19
S41942
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chryso
C;Species: Phanerochaete chrysosporium
C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S44714; S41942
R;Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
Mol. Microbiol. 12, 209-216, 1994
A;Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in th
A;Reference number: S44714; MUID:94335641
A;Accession: S44714
A;Molecule type: mRNA
A;Residues: 1-540 <SI2>
A;Cross-references: EMBL:Z29653; NID:g453222; PIDN:CAA82761.1; PID:g453223
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;479-510/Domain: fungal cellulose-binding domain homology <FCB>


  Query Match             28.4%;  Score 56;  DB 2;  Length 540;
  Best Local Similarity   48.0%;  Pred. No. 48;
  Matches   12;  Conservative    1;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT 27
           |||: ||      | ||| |  | |
```

```
Db    473 GPTVPQW------GQCGGIGYSGST 491


RESULT  20
KSNCLO
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)
N;Alternate names: urishiol oxidase
C;Species: Neurospora crassa
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
C;Accession: A28523; A29762
R;Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
J. Biol. Chem. 263, 885-896, 1988
A;Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
A;Reference number: A28523; MUID:88087214
A;Accession: A28523
A;Molecule type: DNA
A;Residues: 1-619 <GER>
A;Cross-references: EMBL:M14554
R;Germann, U.A.; Lerch, K.
Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986
A;Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospora c
A;Reference number: A29762; MUID:87067412
A;Accession: A29762
A;Molecule type: DNA
A;Residues: 379-619 <GE2>
A;Cross-references: GB:M14554; NID:g168823; PIDN:AAA33590.1; PID:g168824
C;Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
C;Genetics:
A;Introns: 86/3
C;Superfamily: laccase
C;Keywords: copper; glycoprotein; oxidoreductase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-49/Domain: propeptide #status predicted <PRO>
F;50-619/Product: laccase #status predicted <MAT>
F;84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F;216-372/Domain: middle beta-barrel #status predicted <BB2>
F;431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F;139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;144,480/Binding site: copper (His) (type 2) #status predicted
F;146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred
F;477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted


  Query Match             28.4%;  Score 56;  DB 1;  Length 619;
  Best Local Similarity   63.6%;  Pred. No. 55;
  Matches   14;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy      11 AARAGGGCGGGGIEGPTLRQ-W 31
           | | ||| ||||   || || |
Db      44 AERYGGG-GGGGCNSPTNRQCW 64


RESULT  21
KSNCLT
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)
N;Alternate names: urishiol oxidase
C;Species: Neurospora crassa
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
C;Accession: B28523
R;Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
J. Biol. Chem. 263, 885-896, 1988
A;Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
A;Reference number: A28523; MUID:88087214
A;Accession: B28523
A;Molecule type: DNA
A;Residues: 1-619 <GER>
A;Cross-references: EMBL:M18334; NID:g168827; PIDN:AAA33592.1; PID:g168828
C;Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
C;Genetics:
A;Introns: 86/3
C;Superfamily: laccase
C;Keywords: copper; glycoprotein; oxidoreductase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-49/Domain: propeptide #status predicted <PRO>
F;50-619/Product: laccase #status predicted <MAT>
F;84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F;216-372/Domain: middle beta-barrel #status predicted <BB2>
F;431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F;139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predict
F;144,480/Binding site: copper (His) (type 2) #status predicted
F;146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status p
F;477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted


  Query Match             28.4%;  Score 56;  DB 1;  Length 619;
  Best Local Similarity   63.6%;  Pred. No. 55;
  Matches   14;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy      11 AARAGGGCGGGGIEGPTLRQ-W 31
           | | ||| ||||   || || |
Db      44 AERYGGG-GGGGCNSPTNRQCW 64


RESULT  22
E70895
hypothetical glycine-rich protein Rv1087 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C;Accession: E70895
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: E70895
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-767 <COL>
A;Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17203.1; PID:e125
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1087
C;Superfamily: unassigned collagens


  Query Match             28.4%;  Score 56;  DB 2;  Length 767;
  Best Local Similarity   46.7%;  Pred. No. 66;
  Matches   14;  Conservative    0;  Mismatches   12;  Indels    4;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPTLRQWL 32
           |||      |  ||| || |  ||    ||
Db     681 GPTNFGLNGAGGGGGVGGNGATGP----WL 706


RESULT  23
T49530
related to glycine-rich cell wall structural protein [imported] - Neurospora crassa
N;Alternate names: protein B21J21.90
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49530
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
Submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49530
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-180 <SCH>
A;Cross-references: EMBL:AL355929; GSPDB:GN00116; NCSP:B21J21.90
A;Experimental source: BAC clone B21J21; strain OR74A
C;Genetics:
A;Gene: NCSP:B21J21.90
A;Map position: 6
```

```
  Query Match            27.9%;  Score 55;  DB 2;  Length 180;
  Best Local Similarity  73.3%;  Pred. No. 24;
  Matches   11;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      12 ARAGGGCGGGGIEGP 26
           | |||| |||| ||
Db      58 ADAGGGAGGGGGGGP 72


RESULT  24
T49792
hypothetical protein B9J10.290 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49792
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49792
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-201 <SCH>
A;Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.290
A;Experimental source: BAC clone B9J10; strain OR74A
C;Genetics:
A;Gene: NCSP:B9J10.290
A;Map position: 6


  Query Match            27.9%;  Score 55;  DB 2;  Length 201;
  Best Local Similarity  52.4%;  Pred. No. 26;
  Matches   11;  Conservative    2;  Mismatches    4;  Indels    4;  Gaps    1;

Qy      13 RAGGGCGGGGIEGPTLRQWLA 33
           | ||| |||| | |    :| |
Db      74 RGGGGGGGGGVNG----RWSA 90


RESULT  25
C84890
hypothetical protein At2g45420 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84890
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84890
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-257 <STO>
A;Cross-references: GB:AE002093; NID:g2583113; PIDN:AAB82622.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g45420
A;Map position: 2


  Query Match            27.9%;  Score 55;  DB 2;  Length 257;
  Best Local Similarity  81.8%;  Pred. No. 33;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      15 GGGCGGGGIEG 25
           |||||||| |
Db      15 GGGCGGGGSSG 25


RESULT  26
T26807
hypothetical protein Y41C4A.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C;Accession: T26807
R;Steward, C.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20269
A;Accession: T26807
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-331 <WIL>
A;Cross-references: EMBL:AL032627; PIDN:CAB54381.1; CESP:Y41C4A.4a
A;Experimental source: clone Y41C4A
C;Genetics:
A;Gene: CESP:Y41C4A.4a
A;Introns: 24/3; 50/2; 81/3; 159/1; 228/1; 292/3
C;Superfamily: fos/jun DNA-binding domain homology


  Query Match            27.9%;  Score 55;  DB 2;  Length 331;
  Best Local Similarity  69.2%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      15 GGGCGGGGIEGPT 27
           ||| |||| : ||:
Db     167 GGGGGGGGVPGPS 179


RESULT  27
T26808
hypothetical protein Y41C4A.4b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C;Accession: T26808
R;Steward, C.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20269
A;Accession: T26808
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-333 <WIL>
A;Cross-references: EMBL:AL032627; PIDN:CAB54382.1; CESP:Y41C4A.4b
A;Experimental source: clone Y41C4A
C;Genetics:
A;Gene: CESP:Y41C4A.4b
A;Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3
C;Superfamily: fos/jun DNA-binding domain homology


  Query Match            27.9%;  Score 55;  DB 2;  Length 333;
  Best Local Similarity  69.2%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      15 GGGCGGGGIEGPT 27
           ||| |||| : ||:
Db     169 GGGGGGGGVPGPS 181


RESULT  28
T20268
hypothetical protein C56A3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20268
R;Sims, M.
submitted to the EMBL Data Library, July 1996
A;Reference number: Z19244
A;Accession: T20268
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-393 <WIL>
A;Cross-references: EMBL:Z77655; PIDN:CAB01137.1; GSPDB:GN00023; CESP:C56A3.1
A;Experimental source: clone C56A3
C;Genetics:
```

```
A;Gene: CESP:C56A3.1
A;Map position: 5
A;Introns: 51/3; 91/1; 121/1; 331/3


  Query Match           27.9%;  Score 55;  DB 2;  Length 393;
  Best Local Similarity   45.5%;  Pred. No. 47;
  Matches   10;  Conservative    2;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       4 PTLRQWLAARAGGGCGGGGIEG 25
           | ::       |||||||  |
Db      76 PQVQPVYVQSGGGGCGGGGCGG 97


RESULT  29
T47712
MYB transcription factor-like protein - Arabidopsis thaliana
N;Alternate names: protein F1I16.140
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C;Accession: T47712
R;Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24473
A;Accession: T47712
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <BEN>
A;Cross-references: EMBL:AL161667
A;Experimental source: cultivar Columbia; BAC clone F1I16
C;Genetics:
A;Map position: 3
A;Introns: 113/1
A;Note: F1I16.140
C;Superfamily: Arabidopsis myb-related protein 1; myb DNA-binding repeat homology


  Query Match           27.9%;  Score 55;  DB 2;  Length 399;
  Best Local Similarity   55.6%;  Pred. No. 48;
  Matches   10;  Conservative    2;  Mismatches    0;  Indels    6;  Gaps    1;

Qy      15 GGGCGGGG------IEGP 26
           ||||||||      ::||
Db      41 GGGCGGGGGGIRSKVKGP 58


RESULT  30
A32466
numb protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Oct-1992 #sequence_revision 21-Oct-1992 #text_change 24-Sep-1998
C;Accession: A32466
R;Uemura, T.; Shepherd, S.; Ackerman, L.; Jan, L.Y.; Jan, Y.N.
Cell 58, 349-360, 1989
A;Title: numb, a gene required in determination of cell fate during sensory organ format
A;Reference number: A32466; MUID:89324081
A;Accession: A32466
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-556 <UEM>
A;Cross-references: GB:M27815; NID:g158000; PID:g158001
C;Genetics:
A;Gene: FlyBase:numb
A;Cross-references: FlyBase:FBgn0002973


  Query Match           27.9%;  Score 55;  DB 2;  Length 556;
  Best Local Similarity   42.3%;  Pred. No. 64;
  Matches   11;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGGIEGPTLRQWLA 33
           | ||:  |   |||| : |   :|:|
Db     486 QTLASGTGAAVGGGGPDDPFDAEWVA 511


RESULT  31
H81739
alanyl-tRNA synthetase TC0125 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C;Accession: H81739
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: H81739
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-875 <TET>
A;Cross-references: GB:AE002279; GB:AE002160; NID:g7190148; PIDN:AAF39003.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0125
C;Superfamily: alanine--tRNA ligase


  Query Match           27.9%;  Score 55;  DB 2;  Length 875;
  Best Local Similarity   28.6%;  Pred. No. 95;
  Matches   14;  Conservative    5;  Mismatches   16;  Indels   14;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIE-------------GPTLRQWLAAR 35
           |:   | : | |  || |||  :                 ||||:: |
Db     825 IDAQALLEELLAPYGGRCGGKAVSAQGSSKELPQIEVLNTILRQWISTR 873


RESULT  32
F70954
probable lsr2 protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: F70954
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: F70954
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-112 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08947.1; PID:g21139
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: lsr2


  Query Match           27.7%;  Score 54.5;  DB 2;  Length 112;
  Best Local Similarity   33.3%;  Pred. No. 18;
  Matches   13;  Conservative    6;  Mismatches    7;  Indels   13;  Gaps    2;

Qy       6 LRQWLAA----------RAGGGCGGGGI---EGPTLRQW 31
           |:||:||          |:| | | | |   :   :|:|
Db      48 LKQWVAAGRRVGGRRRGRSGSGRGRGAIDREQSAAIREW 86


RESULT  33
A54507
dnaK-type molecular chaperone - fluke (Schistosoma japonicum) (fragment)
N;Alternate names: heat shock protein 70
C;Species: Schistosoma japonicum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 20-Aug-1999
C;Accession: A54507
R;Hedstrom, R.; Culpepper, J.; Schinski, V.; Agabian, N.; Newport, G.
```

```
Mol. Biochem. Parasitol. 29, 275-282, 1988
A;Title: Schistosome heat-shock proteins are immunologically distinct host-like antigens
A;Reference number: A54507; MUID:88318804
A;Accession: A54507
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-198 <HED>
A;Cross-references: GB:M21011; NID:g161022; PIDN:AAA29897.1; PID:g161023
C;Function:
A;Description: involved in protein folding and assembling/disassembling of protein compl
C;Superfamily: heat shock protein 70
C;Keywords: ATP; molecular chaperone


  Query Match             27.7%;  Score 54.5;  DB 2;  Length 198;
  Best Local Similarity   41.4%;  Pred. No. 29;
  Matches   12;  Conservative    3;  Mismatches    3;  Indels   11;  Gaps    1;

Qy      13 RAGG----------GCGGGGIEGPTLRQ 30
           ||||          | |||| :|||: :
Db     168 RAGGGVPSGMPGGMPGAGGGGGKGPTIEE 196


RESULT  34
G02371
U1-snRNP binding protein homolog - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C;Accession: G02371
R;Adams, D.
submitted to the EMBL Data Library, January 1996
A;Reference number: H01131
A;Accession: G02371
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-246 <ADA>
A;Cross-references: EMBL:U44798; NID:g1174216; PIDN:AAA86654.1; PID:g1174217
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotei
F;52-119/Domain: ribonucleoprotein repeat homology <RRM>


  Query Match             27.7%;  Score 54.5;  DB 2;  Length 246;
  Best Local Similarity   38.7%;  Pred. No. 36;
  Matches   12;  Conservative    3;  Mismatches    7;  Indels    9;  Gaps    1;

Qy       5 TLRQWLAARAGGGCGG---------GGIEGP 26
           ||: |:  | ||| ||         || : |
Db     131 TLKGWIPRRLGGGLGGKKESGQLRFGGRDRP 161


RESULT  35
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C;Accession: A56038
R;Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A;Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaste
A;Reference number: A56038; MUID:95021209
A;Accession: A56038
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1028 <GAR>
A;Cross-references: GB:U11383; NID:g520526; PIDN:AAB60216.1; PID:g520527
C;Genetics:
A;Gene: ovo
A;Cross-references: FlyBase:FBgn0003028


  Query Match             27.7%;  Score 54.5;  DB 2;  Length 1028;
  Best Local Similarity   57.9%;  Pred. No. 1.2e+02;
  Matches   11;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    1;

Qy      11 AARAGGGC---GGGGIEGP 26
           |   ||||   ||||  ||
Db      71 AGSGGGGCTGNGGGGASGP 89


RESULT  36
S16356
ovo protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C;Accession: S16356
R;Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A;Title: The ovo gene of Drosophila encodes a zinc finger protein required for female
A;Reference number: S16356; MUID:91293102
A;Accession: S16356
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1213 <MEV>
A;Cross-references: EMBL:X59772
C;Genetics:
A;Gene: FlyBase:ovo
A;Cross-references: FlyBase:FBgn0003028
A;Introns: 931/3; 1152/3


  Query Match             27.7%;  Score 54.5;  DB 2;  Length 1213;
  Best Local Similarity   57.9%;  Pred. No. 1.4e+02;
  Matches   11;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    1;

Qy      11 AARAGGGC---GGGGIEGP 26
           |   ||||   ||||  ||
Db     434 AGSGGGGCTGNGGGGASGP 452


RESULT  37
JQ1094
hypothetical 20.2K protein - tomato ringspot virus
C;Species: tomato ringspot virus
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Oct-1999
C;Accession: JQ1094
R;Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
J. Gen. Virol. 72, 1505-1514, 1991
A;Title: Nucleotide sequence of tomato ringspot virus RNA-2.
A;Reference number: JQ1093; MUID:91311402
A;Accession: JQ1094
A;Status: translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-201 <ROT>
A;Cross-references: GB:D12477; GB:D01129; NID:g222674; PIDN:BAA02044.1; PID:d1002526;
A;Experimental source: strain raspberry


  Query Match             27.4%;  Score 54;  DB 2;  Length 201;
  Best Local Similarity   57.7%;  Pred. No. 34;
  Matches   15;  Conservative    1;  Mismatches    6;  Indels    4;  Gaps    1;

Qy      13 RAGGGCGGGGIE----GPTLRQWLAA 34
           ||||| |||| |    | || : | |
Db      13 RAGGGGGGGGKEVFKAGRTLLKVLKA 38


RESULT  38
A49447
transcription factor Brn-2 - rat
N;Alternate names: class III POU domain protein brain-2
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 20-Feb-1998
C;Accession: A49447
R;Li, P.; He, X.; Gerrero, M.R.; Mok, M.; Aggarwal, A.; Rosenfeld, M.G.
Genes Dev. 7, 2483-2496, 1993
A;Title: Spacing and orientation of bipartite DNA-binding motifs as potential functio
```

```
A;Reference number: A49447; MUID:94102531
A;Accession: A49447
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-445 <LI1>
A;Cross-references: GB:L27663; NID:g443687
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:141696)
C;Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;68-90/Region: glycine-rich
F;125-151/Region: glutamine-rich
F;153-165/Region: histidine/proline-rich
F;213-261/Region: histidine/proline-rich
F;271-338/Domain: POU domain homology <POU>
F;357-413/Domain: homeobox homology <HOX>


  Query Match             27.4%;  Score 54;  DB 1;  Length 445;
  Best Local Similarity   60.0%;  Pred. No. 68;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGG 22
           ||: | : || ||||
Db      60 QWITALSHGGSGGGG 74


RESULT  39
T09084
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: T09084
R;Molendijk, A.J.; Irvine, R.F.
Plant Mol. Biol. 37, 53-66, 1998
A;Title: Inositide signalling in Chlamydomonas: Characterization of a phosphatidylinosit
A;Reference number: Z16411; MUID:98281574
A;Accession: T09084
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-490 <MOL>
A;Cross-references: EMBL:U97663; NID:g2109290; PIDN:AAC50018.1; PID:g2109291
A;Experimental source: strain cw-15
C;Genetics:
A;Introns: 265/3; 331/3; 370/3; 455/1; 481/3


  Query Match             27.4%;  Score 54;  DB 2;  Length 490;
  Best Local Similarity   45.7%;  Pred. No. 74;
  Matches   16;  Conservative    2;  Mismatches    7;  Indels   10;  Gaps    3;

Qy       3 GPTLRQWLAARAGGGCGGGGI---EGPTLR--QWL 32
           ||     | |  ||| ||||    :| | |  :||
Db     231 GP-----LLAAGGGGGGGGSSPGDGSTARWDEWL 260


RESULT  40
E70948
probable amidase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70948
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: E70948
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-495 <COL>
A;Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16640.1; PID:g282
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3175
C;Superfamily: indoleacetamide hydrolase


  Query Match             27.4%;  Score 54;  DB 2;  Length 495;
  Best Local Similarity   55.0%;  Pred. No. 74;
  Matches   11;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       3 GPTLRQWLAARAGGGCGGGG 22
           | |   | |||  ||  |||
Db     147 GRTNNPWDAARTSGGSAGGG 166


RESULT  41
F64408
coenzyme F420 hydrogenase (EC 1.12.99.1) beta chain homolog - Methanococcus jannaschi
N;Alternate names: coenzyme F420-reducing hydrogenase, beta subunit
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: F64408
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A;Reference number: A64300; MUID:96337999
A;Accession: F64408
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-620 <BUL>
A;Cross-references: GB:U67531; GB:L77117; NID:g2826342; PIDN:AAB98876.1; PID:g1591554
C;Genetics:
A;Map position: REV794850-792988
C;Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]
C;Keywords: oxidoreductase
F;488-542/Domain: ferredoxin 2[4Fe-4S] homology <FER>


  Query Match             27.4%;  Score 54;  DB 2;  Length 620;
  Best Local Similarity   43.5%;  Pred. No. 90;
  Matches   10;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 EGPTLRQWLAARAGGGCGGGGIE 24
           ||| :|  ||   || |  | ::
Db     418 EGPLVRATLACPGGGNCSSGLVD 440


RESULT  42
T13807
potassium channel protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13807
R;Goldstein, S.A.; Price, L.A.; Rosenthal, D.N.; Pausch, M.H.
Proc. Natl. Acad. Sci. U.S.A. 93, 13256-13261, 1996
A;Title: ORK1, a potassium-selective leak channel with two pore domains cloned from D
A;Reference number: Z17770; MUID:97075152
A;Accession: T13807
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1001 <GOL>
A;Cross-references: EMBL:U55321; NID:g3808067; PID:g3808068; PIDN:AAC69250.1
C;Genetics:
A;Gene: ORK1
A;Cross-references: FlyBase:FBgn0017561
A;Map position: 1


  Query Match             27.4%;  Score 54;  DB 2;  Length 1001;
```

```
  Best Local Similarity   52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGGIEGPTLRQ 30
           |  || |||  |||||   : :|
Db     761 QQQAAAAGGAAGGGGISRGSRKQ 783


RESULT  43
F72771
probable lysyl-tRNA Synthetase APE0161 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: F72771
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: F72771
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-562 <KAW>
A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79072.1; PID:g5103551
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0161
C;Superfamily: Lyme disease spirochete lysine--tRNA ligase


  Query Match             27.2%;  Score 53.5;  DB 2;  Length 562;
  Best Local Similarity   39.3%;  Pred. No. 94;
  Matches   11;  Conservative    4;  Mismatches   10;  Indels    3;  Gaps    1;

Qy       8 QWLAARAGG---GCGGGGIEGPTLRQWL 32
           :|:: ||||        |  | | |:||
Db     293 EWVSLRAGGREADMSSSGFTGITPREWL 320


RESULT  44
T37948
probable U1 small nuclear ribonucleoprotein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T37948
R;Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21756
A;Accession: T37948
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-261 <SKE>
A;Cross-references: EMBL:Z98974; PIDN:CAB11649.1; GSPDB:GN00066; SPDB:SPAC19A8.13
A;Experimental source: strain 972h-; cosmid c19A8
C;Genetics:
A;Gene: SPDB:SPAC19A8.13
A;Map position: 1
C;Superfamily: transformer-2 sex-determining protein; ribonucleoprotein repeat homology


  Query Match             26.9%;  Score 53;  DB 2;  Length 261;
  Best Local Similarity   50.0%;  Pred. No. 55;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3 GPTLRQWLAARAGGCGG 20
           | |:: ||  : ||| ||
Db     178 GRTVKGWLPRKLGGGLGG 195


RESULT  45
T19389
hypothetical protein C18D11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T19389
R;Barlow, K.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19118
A;Accession: T19389
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-309 <WIL>
A;Cross-references: EMBL:Z92826; PIDN:CAB07322.1; GSPDB:GN00021; CESP:C18D11.4
A;Experimental source: clone C18D11
C;Genetics:
A;Gene: CESP:C18D11.4
A;Map position: 3
A;Introns: 17/3; 39/2; 146/3; 241/2
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro


  Query Match             26.9%;  Score 53;  DB 2;  Length 309;
  Best Local Similarity   52.6%;  Pred. No. 63;
  Matches   10;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 PTLRQWLAARAGGCGGGG 22
           ||  |::  | |||  |||
Db     160 PTPGQYMGDRRGGGSSGGG 178


Search completed: December 26, 2001, 10:28:44
Job time: 156 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:27:54 ; Search time 10.22 Seconds
(without alignments)
129.152 Million cell updates/sec

Title: US-09-422-838C-33
Perfect score: 197
Sequence: 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	63.5	32.2	209	1	NT5_RAT	P34131 rattus norv
2	60.5	30.7	210	1	NT5_HUMAN	P34130 homo sapien
3	60	30.5	497	1	FXD2_HUMAN	O60548 homo sapien
4	58	29.4	875	1	SYA_CHLTR	O84754 chlamydia t
5	57	28.9	339	1	HXD9_MOUSE	P28357 mus musculu
6	56.5	28.7	105	1	INS_BOVIN	P01317 bos taurus
7	56.5	28.7	105	1	INS_SHEEP	P01318 ovis aries
8	56	28.4	619	1	LAC1_NEUCR	P06811 neurospora
9	56	28.4	619	1	LAC2_NEUCR	P10574 neurospora
10	55	27.9	556	1	NUMB_DROME	P16554 drosophila
11	55	27.9	875	1	SYA_CHLMU	Q9plh5 chlamydia m
12	54.5	27.7	112	1	LSR2_MYCTU	O06285 mycobacteri
13	54.5	27.7	198	1	HS70_SCHJA	P12795 schistosoma
14	54.5	27.7	1028	1	OVO_DROME	P51521 drosophila
15	54	27.4	201	1	YR21_TRSVR	P25245 tomato ring
16	54	27.4	620	1	Y870_METJA	Q58280 methanococc
17	54	27.4	1001	1	ORK1_DROME	Q94526 drosophila
18	53.5	27.2	266	1	SCO2_HUMAN	O43819 homo sapien
19	53.5	27.2	562	1	SYK_AERPE	Q9yft9 aeropyrum p
20	53	26.9	497	1	GATA_MYCLE	O33105 mycobacteri
21	53	26.9	716	1	E2BE_RAT	Q64350 rattus norv
22	52.5	26.6	174	1	SSB_RHOSH	Q9zaq8 rhodobacter
23	52.5	26.6	341	1	SPIN_CBEPV	P23061 choristoneu
24	52.5	26.6	370	1	CYB_MICIK	Q9mlk2 micropechis
25	52.5	26.6	448	1	RU17_DROME	P17133 drosophila
26	52.5	26.6	969	1	PAC4_HUMAN	P29122 homo sapien
27	52	26.4	333	1	SIX3_MOUSE	Q62233 mus musculu
28	52	26.4	394	1	FXD3_CHICK	P79772 gallus gall
29	52	26.4	426	1	HKLB_LYCES	O22300 lycopersico
30	52	26.4	443	1	OC3N_HUMAN	P20265 homo sapien
31	52	26.4	445	1	OC3N_MOUSE	P31360 mus musculu
32	52	26.4	448	1	SRF_XENLA	P23790 xenopus lae
33	52	26.4	462	1	ERR1_MOUSE	O08580 mus musculu
34	52	26.4	494	1	GATA_MYCTU	O53258 mycobacteri
35	52	26.4	545	1	CG12_YEAST	P20438 saccharomyc
36	52	26.4	546	1	CG11_YEAST	P20437 saccharomyc
37	52	26.4	584	1	CNA1_DROME	P12252 drosophila
38	52	26.4	593	1	K1CJ_HUMAN	P13645 homo sapien
39	52	26.4	1168	1	MYSC_ACACA	P10569 acanthamoeb
40	52	26.4	1176	1	NIR_NEUCR	P38681 neurospora
41	52	26.4	1178	1	PHYB_SORBI	P93527 sorghum bic
42	51.5	26.1	378	1	RU17_MOUSE	Q62376 mus musculu
43	51.5	26.1	437	1	RU17_HUMAN	P08621 homo sapien
44	51.5	26.1	440	1	FXGA_CHICK	Q98937 gallus gall
45	51.5	26.1	471	1	RU17_XENLA	P09406 xenopus lae
46	51.5	26.1	864	1	KLTK_HUMAN	P29376 homo sapien
47	51	25.9	323	1	JUND_CHICK	P27921 gallus gall
48	51	25.9	348	1	SXL_CERCA	O61374 ceratitis c
49	51	25.9	440	1	DCO_DROME	O76324 drosophila
50	51	25.9	475	1	EVX2_MOUSE	P49749 mus musculu
51	51	25.9	504	1	ATIN_HSVBP	P30020 bovine herp
52	51	25.9	569	1	K1CJ_MOUSE	P02535 mus musculu
53	51	25.9	702	1	TBX2_HUMAN	Q13207 homo sapien
54	51	25.9	888	1	KLTK_MOUSE	P08923 mus musculu
55	51	25.9	1043	1	FTF1_DROME	P33244 drosophila
56	51	25.9	1250	1	TP3A_DROME	Q9ng98 drosophila
57	51	25.9	1322	1	SUS_DROME	P22293 drosophila
58	51	25.9	1454	1	KDGE_DROME	Q09103 drosophila
59	50.5	25.6	312	1	TRPF_CRYNE	P27710 cryptococcu
60	50.5	25.6	391	1	SOX1_MOUSE	P53783 mus musculu
61	50.5	25.6	427	1	AROA_AERPE	Q9yek9 aeropyrum p
62	50.5	25.6	608	1	OM70_HUMAN	O94826 homo sapien
63	50.5	25.6	757	1	CIKF_HUMAN	Q14003 homo sapien
64	50.5	25.6	769	1	CIKF_MOUSE	Q63959 mus musculu
65	50.5	25.6	889	1	CIKF_RAT	Q01956 rattus norv
66	50	25.4	205	1	YJ11_MYCTU	O07722 mycobacteri
67	50	25.4	297	1	XERC_MYCLE	Q9cbu0 mycobacteri
68	50	25.4	367	1	BET3_MESAU	O09029 mesocricetu
69	50	25.4	377	1	DNAJ_LISMO	Q9s5a3 listeria mo
70	50	25.4	401	1	HB9_HUMAN	P50219 homo sapien
71	50	25.4	427	1	RU17_ARATH	Q42404 arabidopsis
72	50	25.4	466	1	HN3A_RAT	P23512 rattus norv
73	50	25.4	468	1	HN3A_MOUSE	P35582 mus musculu
74	50	25.4	485	1	ONC2_HUMAN	O95948 homo sapien
75	50	25.4	584	1	RECN_SYNY3	P74374 synechocyst
76	50	25.4	757	1	ECR_LUCCU	O18531 lucilia cup
77	50	25.4	904	1	DPO1_MYCTU	Q07700 mycobacteri
78	50	25.4	1264	1	CYA5_RABIT	P40144 oryctolagus
79	50	25.4	4499	1	DYHA_CHLRE	Q39610 chlamydomon
80	49.5	25.1	112	1	LSR2_MYCLE	P24094 mycobacteri
81	49.5	25.1	333	1	CBBR_XANFL	P25545 xanthobacte
82	49.5	25.1	342	1	HXD9_HUMAN	P28356 homo sapien
83	49.5	25.1	651	1	HS70_ONCMY	P08108 oncorhynchu
84	49	24.9	104	1	HOL3_HOLDI	Q25055 holotrichia
85	49	24.9	248	1	RASH_RRASV	P01114 rasheed rat
86	49	24.9	323	1	HXDB_MOUSE	P23813 mus musculu
87	49	24.9	353	1	ROD_RAT	Q9jj54 rattus norv
88	49	24.9	385	1	RO32_XENLA	P51992 xenopus lae
89	49	24.9	392	1	HME1_HUMAN	Q05925 homo sapien
90	49	24.9	444	1	GAT6_MOUSE	Q61169 mus musculu
91	49	24.9	445	1	OC3N_RAT	P56222 rattus norv
92	49	24.9	476	1	EVX2_HUMAN	Q03828 homo sapien
93	49	24.9	495	1	BRN1_MOUSE	P31361 mus musculu
94	49	24.9	497	1	BRN1_RAT	Q63262 rattus norv
95	49	24.9	500	1	BRN1_HUMAN	P20264 homo sapien
96	49	24.9	513	1	GUX1_TRIRE	P00725 trichoderma
97	49	24.9	517	1	Y967_TREPA	O83933 treponema p
98	49	24.9	546	1	PGMU_ECOLI	P36938 escherichia
99	49	24.9	631	1	YCIQ_ECOLI	P45848 escherichia
100	49	24.9	634	1	HS70_CHICK	P08106 gallus gall

ALIGNMENTS

RESULT 1

```
NT5_RAT
ID   NT5_RAT        STANDARD;      PRT;   209 AA.
AC   P34131;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROTROPHIN-5 PRECURSOR (NT-5) (NEUTROPHIC FACTOR 5) (NEUROTROPHIN-4)
DE   (NT-4) (NEUTROPHIC FACTOR 4).
GN   NTF5 OR NTF4 OR NT4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92212967; PubMed=1313578;
RA   Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
RA   Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
RA   Yancopoulos G.D.;
RT   "Mammalian neurotrophin-4: structure, chromosomal localization,
RT   tissue distribution, and receptor specificity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92075279; PubMed=1742028;
RA   Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
RA   Rosenthal A.;
RT   "Neurotrophin-5: a novel neurotrophic factor that activates trk and
RT   trkB.";
RL   Neuron 7:857-866(1991).
CC   -!- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR
CC       SENSORY AND SYMPATHETIC NEURONS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,
CC       HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT
CC       TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86742; AAA41728.1; -.
DR   EMBL; S69323; AAB20548.1; -.
DR   PIR; JH0504; JH0504.
DR   PIR; B42687; B42687.
DR   HSSP; P23560; 1BND.
DR   InterPro; IPR002072; NGF.
DR   Pfam; PF00243; NGF; 1.
DR   PRINTS; PR00268; NGF.
DR   ProDom; PD002052; NGF; 1.
DR   SMART; SM00140; NGF; 1.
DR   PROSITE; PS00248; NGF_1; 1.
DR   PROSITE; PS50270; NGF_2; 1.
KW   Growth factor; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?     79
FT   CHAIN        80    209       NEUROTROPHIN-5.
FT   DISULFID     96    169       BY SIMILARITY.
FT   DISULFID    140    198       BY SIMILARITY.
FT   DISULFID    157    200       BY SIMILARITY.
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    177    177       R -> P (IN REF. 2).
SQ   SEQUENCE   209 AA;  22332 MW;  DF5112C05C5D5B85 CRC64;


  Query Match             32.2%;  Score 63.5;  DB 1;  Length 209;
  Best Local Similarity   37.5%;  Pred. No. 2.2;
  Matches   15;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    1;

Qy        3 GPTLRQWL---------AARAGGGCGGGGIEGPTLRQWLA 33
            |  |||:           |    | | ||||  |   | ||:
Db      128 GSPLRQYFFETRCKAESAGEGGPGVGGGGCRGVDRRHWLS 167


RESULT  2
NT5_HUMAN
ID   NT5_HUMAN      STANDARD;      PRT;   210 AA.
AC   P34130;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROTROPHIN-5 PRECURSOR (NT-5) (NEUTROPHIC FACTOR 5) (NEUROTROPHIN-4)
DE   (NT-4) (NEUTROPHIC FACTOR 4).
GN   NTF5 OR NTF4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=92212967; PubMed=1313578;
RA   Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
RA   Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
RA   Yancopoulos G.D.;
RT   "Mammalian neurotrophin-4: structure, chromosomal localization,
RT   tissue distribution, and receptor specificity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92075279; PubMed=1742028;
RA   Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
RA   Rosenthal A.;
RT   "Neurotrophin-5: a novel neurotrophic factor that activates trk and
RT   trkB.";
RL   Neuron 7:857-866(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
RX   MEDLINE=20095835; PubMed=10631974;
RA   Robinson R.C., Radziejewski C., Spraggon G., Greenwald J.,
RA   Kostura M.R., Burtnick L.D., Stuart D.I., Choe S., Jones E.Y.;
RT   "The structures of the neurotrophin 4 homodimer and the brain-derived
RT   neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-
RT   binding site.";
RL   Protein Sci. 8:2589-2597(1999).
CC   -!- FUNCTION: TARGET-DERIVED SURVIVAL FACTOR FOR PERIPHERAL SENSORY
CC       SYMPATHETIC NEURONS.
CC   -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN PROSTATE, LOWER LEVELS
CC       IN THYMUS, PLACENTA, AND SKELETAL MUSCLE. EXPRESSED IN EMBRYONIC
CC       AND ADULT TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86528; AAA60154.1; -.
DR   PIR; JH0503; JH0503.
DR   PIR; A42687; A42687.
DR   PDB; 1B8M; 09-FEB-99.
DR   PDB; 1B98; 26-FEB-99.
DR   MIM; 162662; -.
DR   InterPro; IPR002072; NGF.
DR   Pfam; PF00243; NGF; 1.
DR   PRINTS; PR00268; NGF.
DR   ProDom; PD002052; NGF; 1.
DR   SMART; SM00140; NGF; 1.
DR   PROSITE; PS00248; NGF_1; 1.
```

```
DR   PROSITE; PS50270; NGF_2; 1.
KW   Growth factor; Signal; 3D-structure.
FT   SIGNAL        1     24       POTENTIAL.
FT   PROPEP       25     80
FT   CHAIN        81    210       NEUROTROPHIN-5.
FT   DISULFID     97    170
FT   DISULFID    141    199
FT   DISULFID    158    201
FT   CARBOHYD     76     76       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   210 AA;  22426 MW;  DBC6A30195E139AD CRC64;


  Query Match             30.7%;  Score 60.5;  DB 1;  Length 210;
  Best Local Similarity   35.0%;  Pred. No. 4.5;
  Matches   14;  Conservative    3;  Mismatches   14;  Indels    9;  Gaps    1;

Qy       3 GPTLRQWL---------AARAGGGCGGGGIEGPTLRQWLA 33
           |  |||:           |   | | ||||  |   | |::
Db     129 GSPLRQYFFETRCKADNAEEGGPGAGGGGCRGVDRRHWVS 168


RESULT   3
FXD2_HUMAN
ID   FXD2_HUMAN     STANDARD;      PRT;   497 AA.
AC   O60548;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FORKHEAD BOX PROTEIN D2 (FORKHEAD-RELATED PROTEIN FKHL17) (FORKHEAD-
DE   RELATED TRANSCRIPTION FACTOR 9) (FREAC-9).
GN   FOXD2 OR FKHL17 OR FREAC9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98066765; PubMed=9403061;
RA   Ernstsson S., Betz R., Lagercrantz S., Larsson C., Ericksson S.,
RA   Cederberg A., Carlsson P., Enerbaeck S.;
RT   "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
RT   expressed human forkhead gene that maps to chromosome 1p32-p34.";
RL   Genomics 46:78-85(1997).
RN   [2]
RP   REVISIONS.
RA   Enerbaeck S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF042832; AAC15421.1; -.
DR   HSSP; Q63245; 2HFH.
DR   MIM; 602211; -.
DR   InterPro; IPR001766; Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation.
FT   DOMAIN       90     94       POLY-ALA.
FT   DOMAIN      101    104       POLY-ALA.
FT   DNA_BIND    126    217       FORK-HEAD.
FT   DOMAIN      247    250       POLY-ALA.
FT   DOMAIN      296    306       POLY-ALA.
FT   DOMAIN      398    409       POLY-GLY.
FT   DOMAIN      421    426       POLY-GLY.
FT   DOMAIN      442    445       POLY-ALA.
SQ   SEQUENCE   497 AA;  49007 MW;  EAAF498D216BE019 CRC64;


  Query Match             30.5%;  Score 60;  DB 1;  Length 497;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches   14;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy       4 PT--LRQWLAARAGGGCGGGG 22
           ||  |||  |   |||| ||||
Db     385 PTALLRQGLKTDAGGGAGGGG 405


RESULT   4
SYA_CHLTR
ID   SYA_CHLTR     STANDARD;      PRT;   875 AA.
AC   O84754;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
GN   ALAS OR CT749.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ALANINE + TRNA(ALA) = AMP +
CC       PYROPHOSPHATE + L-ALANYL-TRNA(ALA).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001346; AAC68344.2; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR003156; DHHA1.
DR   InterPro; IPR002318; tRNA-synt_2c.
DR   Pfam; PF02272; DHHA1; 1.
DR   Pfam; PF01411; tRNA-synt_2c; 2.
DR   PRINTS; PR00980; TRNASYNTHALA.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
SQ   SEQUENCE   875 AA;  97671 MW;  81C2DA7B29A5D11D CRC64;


  Query Match             29.4%;  Score 58;  DB 1;  Length 875;
  Best Local Similarity   30.6%;  Pred. No. 28;
  Matches   15;  Conservative    5;  Mismatches   15;  Indels   14;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIE-------------GPTLRQWLAAR 35
           ::  ||   | |  || |||  |               |||||:: :
```

```
Db    825 VQAHTLLAELLAPYGGRCGGKAISAQGSSAELPQIEFLNKTLRQWISTQ 873


RESULT   5
HXD9_MOUSE
ID   HXD9_MOUSE     STANDARD;      PRT;   339 AA.
AC   P28357;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HOMEOBOX PROTEIN HOX-D9 (HOX-4.4) (HOX-5.2).
GN   HOXD9 OR HOXD-9 OR HOX-4.4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92224884; PubMed=1348690;
RA   Renucci A.G.P., Zappavigna V., Zakany J., Izpisua-Belmonte J.-C.,
RA   Buerki K., Douboule D.;
RT   "Comparison of mouse and human HOX-4 complexes defines conserved
RT   sequences involved in the regulation of Hox-4.4.";
RL   EMBO J. 11:1459-1468(1992).
RN   [2]
RP   SEQUENCE OF 272-331 FROM N.A.
RX   MEDLINE=89356622; PubMed=2569970;
RA   Dolle P., Duboule D.;
RT   "Two gene members of the murine HOX-5 complex show regional and cell-
RT   type specific expression in developing limbs and gonads.";
RL   EMBO J. 8:1507-1515(1989).
CC   -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC       A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC       SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LIMB BUDS.
CC   -!- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEOBOX PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62669; CAA44542.1; -.
DR   EMBL; X14714; CAB57813.1; -.
DR   PIR; S09398; S09398.
DR   PIR; S09569; S09569.
DR   PIR; S20880; S20880.
DR   HSSP; P02833; 1SAN.
DR   TRANSFAC; T01755; -.
DR   MGD; MGI:96210; Hoxd9.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW   Transcription regulation.
FT   DOMAIN      113    140       GLY-RICH.
FT   DOMAIN      119    131       POLY-GLY.
FT   DOMAIN      163    176       SER/THR-RICH.
FT   DNA_BIND    272    331       HOMEOBOX.
SQ   SEQUENCE   339 AA;  34992 MW;  370DC47C6929F7E1 CRC64;


  Query Match             28.9%;  Score 57;  DB 1;  Length 339;
  Best Local Similarity   40.6%;  Pred. No. 16;
  Matches   13;  Conservative    2;  Mismatches    9;  Indels    8;  Gaps    1;

Qy       3 GPTLRQWL--------AARAGGGCGGGGIEGP 26
           |  :| |:          |  ||| |||| ||
Db     101 GRYVRSWMEPLPGFPGGAGGGGGSGGGGGGGP 132


RESULT   6
INS_BOVIN
ID   INS_BOVIN      STANDARD;      PRT;   105 AA.
AC   P01317;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   INSULIN PRECURSOR.
GN   INS.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88288209; PubMed=2456452;
RA   D'Agostino J., Younes M.A., White J.W., Besch P.K., Field J.B.,
RA   Frazier M.L.;
RT   "Cloning and nucleotide sequence analysis of complementary
RT   deoxyribonucleic acid for bovine preproinsulin.";
RL   Mol. Endocrinol. 1:327-331(1987).
RN   [2]
RP   SEQUENCE OF 25-105.
RX   MEDLINE=71166442; PubMed=4928892;
RA   Nolan C., Margoliash E., Peterson J.D., Steiner D.F.;
RT   "The structure of bovine proinsulin.";
RL   J. Biol. Chem. 246:2780-2795(1971).
RN   [3]
RP   SEQUENCE OF 25-54.
RA   Sanger F., Tuppy H.;
RT   "The amino-acid sequence in the phenylalanyl chain of insulin. 2. The
RT   investigation of peptides from enzymic hydrolysates.";
RL   Biochem. J. 49:481-490(1951).
RN   [4]
RP   SEQUENCE OF 57-82.
RX   MEDLINE=71116409; PubMed=5545080;
RA   Steiner D.F., Cho S., Oyer P.E., Terris S., Peterson J.D.,
RA   Rubenstein A.H.;
RT   "Isolation and characterization of proinsulin C-peptide from bovine
RT   pancreas.";
RL   J. Biol. Chem. 246:1365-1374(1971).
RN   [5]
RP   SEQUENCE OF 57-82.
RX   MEDLINE=71257721; PubMed=5105368;
RA   Salokangas A., Smyth D.G., Markussen J., Sundby F.;
RT   "Bovine proinsulin: amino acid sequence of the C-peptide isolated
RT   from pancreas.";
RL   Eur. J. Biochem. 20:183-189(1971).
RN   [6]
RP   SEQUENCE OF 85-105.
RA   Sanger F., Thompson E.O.P.;
RT   "The amino-acid sequence in the glycyl chain of insulin. 2. The
RT   investigation of peptides from enzymic hydrolysates.";
RL   Biochem. J. 53:366-374(1953).
RN   [7]
RP   AMIDES, SEQUENCE OF 25-54 AND 85-105, AND DISULFIDE BONDS.
RA   Ryle A.P., Sanger F., Smith L.F., Kitai R.;
RT   "The disulphide bonds of insulin.";
RL   Biochem. J. 60:541-556(1955).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY.
RA   Smith G.D., Duax W.L., Dodson E.J., Dodson G.G., de Graaf R.A.G.,
RA   Reynolds C.D.;
RT   "The structure of des-Phe b1 bovine insulin.";
RL   Acta Crystallogr. B 38:3028-3032(1982).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
```

```
RX   MEDLINE=97285914; PubMed=9141131;
RA   Brange J., Dodson G.G., Edwards D.J., Holden P.H., Whittingham J.L.;
RT   "A model of insulin fibrils derived from the X-ray crystal structure
RT   of a monomeric insulin (despentapeptide insulin).";
RL   Proteins 27:507-516(1997).
CC   -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC       INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC       FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC       CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 9 of April 2001;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt009.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M54979; AAA30722.1; -.
DR   PIR; A01585; IPBO.
DR   PIR; A40909; A40909.
DR   PDB; 2INS; 31-MAY-84.
DR   PDB; 1APH; 31-OCT-93.
DR   PDB; 1BPH; 31-OCT-93.
DR   PDB; 1CPH; 31-OCT-93.
DR   PDB; 1DPH; 31-OCT-93.
DR   PDB; 1PID; 07-DEC-96.
DR   InterPro; IPR000739; Insulin_IGF_relaxin.
DR   Pfam; PF00049; Insulin; 1.
DR   PRINTS; PR00276; INSULINA.
DR   PRINTS; PR00277; INSULINB.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT   SIGNAL       1     24
FT   CHAIN       25     54       INSULIN B CHAIN.
FT   PROPEP      57     82       C PEPTIDE.
FT   CHAIN       85    105       INSULIN A CHAIN.
FT   DISULFID    31     91       INTERCHAIN.
FT   DISULFID    43    104       INTERCHAIN.
FT   DISULFID    90     95
FT   TURN        32     32
FT   HELIX       33     46
FT   STRAND      48     48
FT   HELIX       86     90
FT   TURN        91     94
FT   HELIX       97    101
FT   TURN       102    103
FT   STRAND     104    104
SQ   SEQUENCE   105 AA;  11393 MW;  75307CF78E61C06A CRC64;


  Query Match             28.7%;  Score 56.5;  DB 1;  Length 105;
  Best Local Similarity   50.0%;  Pred. No. 6.6;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGP 26
           :|||   |  |    || | ||:|||
Db      58 VEGP---QVGALELAGGPGAGGLEGP 80


RESULT   7
INS_SHEEP
ID   INS_SHEEP      STANDARD;      PRT;   105 AA.
AC   P01318;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   INSULIN PRECURSOR.
GN   INS.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94280618; PubMed=8011164;
RA   Ohlsen S.M., Lugenbeel K.A., Wong E.A.;
RT   "Characterization of the linked ovine insulin and insulin-like growth
RT   factor-II genes.";
RL   DNA Cell Biol. 13:377-388(1994).
RN   [2]
RP   SEQUENCE OF 25-54 AND 85-105.
RA   Brown H., Sanger F., Kitai R.;
RT   "The structure of pig and sheep insulins.";
RL   Biochem. J. 60:556-565(1955).
RN   [3]
RP   SEQUENCE OF 57-82.
RX   MEDLINE=72258016; PubMed=4626369;
RA   Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT   "Determination of the amino acid sequence of the monkey, sheep, and
RT   dog proinsulin C-peptides by a semi-micro Edman degradation
RT   procedure.";
RL   J. Biol. Chem. 247:4866-4871(1972).
CC   -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC       INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC       FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC       CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00659; AAB60625.1; -.
DR   PIR; S16430; INSH.
DR   HSSP; P01315; 9INS.
DR   InterPro; IPR000739; Insulin_IGF_relaxin.
DR   Pfam; PF00049; Insulin; 1.
DR   PRINTS; PR00276; INSULINA.
DR   PRINTS; PR00277; INSULINB.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Glucose metabolism; Signal.
FT   SIGNAL       1     24
FT   CHAIN       25     54       INSULIN B CHAIN.
FT   PROPEP      57     82       C PEPTIDE.
FT   CHAIN       85    105       INSULIN A CHAIN.
FT   DISULFID    31     91       INTERCHAIN.
FT   DISULFID    43    104       INTERCHAIN.
FT   DISULFID    90     95
SQ   SEQUENCE   105 AA;  11235 MW;  8B27C7FB9922BC7A CRC64;


  Query Match             28.7%;  Score 56.5;  DB 1;  Length 105;
  Best Local Similarity   50.0%;  Pred. No. 6.6;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGP 26
           :|||   |  |    || | ||:|||
Db      58 VEGP---QVGALELAGGPGAGGLEGP 80
```

```
RESULT   8
LAC1_NEUCR
ID   LAC1_NEUCR     STANDARD;      PRT;   619 AA.
AC   P06811;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LACCASE PRECURSOR (EC 1.10.3.2) (BENZENEDIOL:OXYGEN OXIDOREDUCTASE)
DE   (URISHIOL OXIDASE) (LACCASE ALLELE OR).
GN   LACC.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=88087214; PubMed=2961749;
RA   Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT   "Characterization of two allelic forms of Neurospora crassa laccase.
RT   Amino- and carboxyl-terminal processing of a precursor.";
RL   J. Biol. Chem. 263:885-896(1988).
RN   [2]
RP   SEQUENCE OF 379-619 FROM N.A.
RX   MEDLINE=87067412; PubMed=2947240;
RA   Germann U.A., Lerch K.;
RT   "Isolation and partial nucleotide sequence of the laccase gene from
RT   Neurospora crassa: amino acid sequence homology of the protein to
RT   human ceruloplasmin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).
CC   -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC       PRODUCTS (PROBABLE).
CC   -!- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE +
CC       2 H(2)O.
CC   -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC       CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC       3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC   -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14554; AAA33590.1; -.
DR   EMBL; M18333; AAA33591.1; -.
DR   PIR; A28523; KSNCLO.
DR   PIR; A29762; A29762.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR002355; MultiCu_oxidse2.
DR   Pfam; PF00394; Cu-oxidase; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR   PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
KW   Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
KW   Glycoprotein; Repeat.
FT   SIGNAL        1     21       POTENTIAL.
FT   PROPEP       22     49
FT   CHAIN        50    606       LACCASE.
FT   PROPEP      607    619
FT   DOMAIN       84    207       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      216    373       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      431    566       PLASTOCYANIN-LIKE 3.
FT   METAL       144    144       COPPER (TYPE 2) (PROBABLE).
FT   METAL       146    146       COPPER (TYPE 3) (PROBABLE).
FT   METAL       189    189       COPPER (TYPE 3) (PROBABLE).
FT   METAL       191    191       COPPER (TYPE 3) (PROBABLE).
FT   METAL       477    477       COPPER (TYPE 1) (PROBABLE).
FT   METAL       480    480       COPPER (TYPE 2) (PROBABLE).
FT   METAL       482    482       COPPER (TYPE 3) (PROBABLE).
FT   METAL       548    548       COPPER (TYPE 3) (PROBABLE).
FT   METAL       549    549       COPPER (TYPE 1) (PROBABLE).
FT   METAL       550    550       COPPER (TYPE 3) (PROBABLE).
FT   METAL       554    554       COPPER (TYPE 1) (PROBABLE).
FT   METAL       559    559       COPPER (TYPE 1) (PROBABLE).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    282    282       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    422    422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   619 AA;  68198 MW;  FDED6D78B65048E3 CRC64;


  Query Match             28.4%;  Score 56;  DB 1;  Length 619;
  Best Local Similarity   63.6%;  Pred. No. 34;
  Matches   14;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy      11 AARAGGGCGGGGIEGPTLRQ-W 31
           | | ||| |||| || || | |
Db      44 AERYGGG-GGGGCNSPTNRQCW 64


RESULT   9
LAC2_NEUCR
ID   LAC2_NEUCR     STANDARD;      PRT;   619 AA.
AC   P10574;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LACCASE PRECURSOR (EC 1.10.3.2) (BENZENEDIOL:OXYGEN OXIDOREDUCTASE)
DE   (URISHIOL OXIDASE) (LACCASE ALLELE TS).
GN   LACC.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88087214; PubMed=2961749;
RA   Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT   "Characterization of two allelic forms of Neurospora crassa laccase.
RT   Amino- and carboxyl-terminal processing of a precursor.";
RL   J. Biol. Chem. 263:885-896(1988).
CC   -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC       PRODUCTS (PROBABLE).
CC   -!- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE +
CC       2 H(2)O.
CC   -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC       CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC       3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC   -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18334; AAA33592.1; -.
DR   PIR; B28523; KSNCLT.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR002355; MultiCu_oxidse2.
DR   Pfam; PF00394; Cu-oxidase; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR   PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
KW   Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
```

```
KW   Glycoprotein; Repeat.
FT   SIGNAL       1     21       POTENTIAL.
FT   PROPEP      22     49
FT   CHAIN       50    606       LACCASE.
FT   PROPEP     607    619
FT   DOMAIN      84    207       PLASTOCYANIN-LIKE 1.
FT   DOMAIN     216    373       PLASTOCYANIN-LIKE 2.
FT   DOMAIN     431    566       PLASTOCYANIN-LIKE 3.
FT   METAL      144    144       COPPER (TYPE 2) (PROBABLE).
FT   METAL      146    146       COPPER (TYPE 3) (PROBABLE).
FT   METAL      189    189       COPPER (TYPE 3) (PROBABLE).
FT   METAL      191    191       COPPER (TYPE 3) (PROBABLE).
FT   METAL      477    477       COPPER (TYPE 1) (PROBABLE).
FT   METAL      480    480       COPPER (TYPE 2) (PROBABLE).
FT   METAL      482    482       COPPER (TYPE 3) (PROBABLE).
FT   METAL      548    548       COPPER (TYPE 3) (PROBABLE).
FT   METAL      549    549       COPPER (TYPE 1) (PROBABLE).
FT   METAL      550    550       COPPER (TYPE 3) (PROBABLE).
FT   METAL      554    554       COPPER (TYPE 1) (PROBABLE).
FT   METAL      559    559       COPPER (TYPE 1) (PROBABLE).
FT   CARBOHYD   139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   282    282       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   422    422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   444    444       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   619 AA;  68120 MW;  0BB6CCDE18841145 CRC64;


  Query Match             28.4%;  Score 56;  DB 1;  Length 619;
  Best Local Similarity   63.6%;  Pred. No. 34;
  Matches   14;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy      11 AARAGGGCGGGGIEGPTLRQ-W 31
           | | ||| ||||   || || |
Db      44 AERYGGG-GGGGCNSPTNRQCW 64


RESULT  10
NUMB_DROME
ID   NUMB_DROME     STANDARD;      PRT;   556 AA.
AC   P16554;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NUMB PROTEIN.
GN   NUMB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89324081; PubMed=2752427;
RA   Uemura T., Shepherd S., Ackerman L., Jan L.Y., Jan Y.N.;
RT   "Numb, a gene required in determination of cell fate during sensory
RT   organ formation in Drosophila embryos.";
RL   Cell 58:349-360(1989).
RN   [2]
RP   STRUCTURE BY NMR OF 64-210.
RX   MEDLINE=99061335; PubMed=9846878;
RA   Li S.-C., Zwahlen C., Vincent S.J., McGlade C.J., Kay L.E., Pawson T.,
RA   Forman-Kay J.D.;
RT   "Structure of a Numb PTB domain-peptide complex suggests a basis for
RT   diverse binding specificity.";
RL   Nat. Struct. Biol. 5:1075-1083(1998).
CC   -!- FUNCTION: NUMB IS REQUIRED IN DETERMINATION OF CELL FATE DURING
CC       SENSORY ORGAN FORMATION IN DROSOPHILA EMBRYOS. IT FUNCTIONS IN
CC       NUCLEI AND SEEMS TO INTERACT WITH NUCLEIC ACIDS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27815; AAA28730.1; -.
DR   PIR; A32466; A32466.
DR   PDB; 2NMB; 04-NOV-98.
DR   FlyBase; FBgn0002973; numb.
DR   InterPro; IPR000050; PID_domain.
DR   Pfam; PF00640; PID; 1.
DR   SMART; SM00462; PTB; 1.
DR   PROSITE; PS01179; PID; 1.
KW   Nuclear protein; ATP-binding; Alternative initiation; 3D-structure.
FT   CHAIN        1    556       NUMB PROTEIN, ZYGOTIC ISOFORM.
FT   CHAIN       42    556       NUMB PROTEIN, MATERNAL ISOFORM
FT                               (PROBABLE).
FT   INIT_MET    42     42       FOR MATERNAL ISOFORM.
FT   NP_BIND     22     29       ATP (POTENTIAL).
FT   DOMAIN      25     57       ARG/LYS-RICH (BASIC).
FT   DOMAIN      81    208       PID.
SQ   SEQUENCE   556 AA;  60628 MW;  4FECAAE9C98FEE71 CRC64;


  Query Match             27.9%;  Score 55;  DB 1;  Length 556;
  Best Local Similarity   42.3%;  Pred. No. 40;
  Matches   11;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGGIEGPTLRQWLA 33
           | ||:  |   |||| : |   :|:|
Db     486 QTLASGTGAAVGGGGPDDPFDAEWVA 511


RESULT  11
SYA_CHLMU
ID   SYA_CHLMU      STANDARD;      PRT;   875 AA.
AC   Q9PLH5;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
GN   ALAS OR TC0125.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ALANINE + TRNA(ALA) = AMP +
CC       PYROPHOSPHATE + L-ALANYL-TRNA(ALA).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE002279; AAF39003.1; -.
DR   TIGR; TC0125; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR003156; DHHA1.
DR   InterPro; IPR002318; tRNA-synt_2c.
DR   Pfam; PF01411; tRNA-synt_2c; 1.
DR   Pfam; PF02272; DHHA1; 1.
DR   PRINTS; PR00980; TRNASYNTHALA.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
SQ   SEQUENCE   875 AA;  98185 MW;  A75F8977A23DC41D CRC64;


  Query Match             27.9%;  Score 55;  DB 1;  Length 875;
  Best Local Similarity   28.6%;  Pred. No. 58;
  Matches   14;  Conservative    5;  Mismatches   16;  Indels   14;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIE--------------GPTLRQWLAAR 35
           |:   | : | |  || |||  :                 ||||:: |
Db     825 IDAQALLEELLAPYGGRCGGKAVSAQGSSKELPQIEVLNTILRQWISTR 873


RESULT  12
LSR2_MYCTU
ID   LSR2_MYCTU     STANDARD;      PRT;   112 AA.
AC   O06285;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LSR2 PROTEIN PRECURSOR.
GN   LSR2 OR RV3597C OR MT3704 OR MTCY07H7B.25.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
CC       LYMPHOPROLIFERATION (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95557; CAB08947.1; -.
DR   EMBL; AE007170; AAK48061.1; -.
DR   TIGR; MT3704; -.
DR   TubercuList; Rv3597c; -.
KW   Antigen; Immune response; Signal; Complete proteome.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    112       LSR2 PROTEIN.
SQ   SEQUENCE   112 AA;  12098 MW;  A4B32E478CBAC3E4 CRC64;


  Query Match             27.7%;  Score 54.5;  DB 1;  Length 112;
  Best Local Similarity   33.3%;  Pred. No. 11;
  Matches   13;  Conservative    6;  Mismatches    7;  Indels   13;  Gaps    2;

Qy       6 LRQWLAA----------RAGGGCGGGGI---EGPTLRQW 31
           |:|||:||          |:| | | | |   :   :|:|
Db      48 LKQWVAAGRRVGGRRRGRSGSGRGRGAIDREQSAAIREW 86


RESULT  13
HS70_SCHJA
ID   HS70_SCHJA     STANDARD;      PRT;   198 AA.
AC   P12795;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-JUL-1993 (Rel. 26, Last annotation update)
DE   HEAT SHOCK 70 KDA PROTEIN (HSP70) (FRAGMENT).
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC   Rhabditophora; Eulecithophora; Revertospermata; Mediofusata;
OC   Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
OC   Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6182;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88318804; PubMed=2457805;
RA   Hedstrom R., Culpepper J., Schinski V., Agabian N., Newport G.;
RT   "Schistosome heat-shock proteins are immunologically distinct
RT   host-like antigens.";
RL   Mol. Biochem. Parasitol. 29:275-282(1988).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21011; AAA29897.1; -.
DR   PIR; A54507; A54507.
DR   HSSP; P02593; 1CMF.
DR   InterPro; IPR001023; HSP70.
DR   PROSITE; PS00297; HSP70_1; PARTIAL.
DR   PROSITE; PS00329; HSP70_2; PARTIAL.
DR   PROSITE; PS01036; HSP70_3; PARTIAL.
KW   ATP-binding; Heat shock.
FT   NON_TER       1      1
SQ   SEQUENCE   198 AA;  21845 MW;  800F8586046D5313 CRC64;


  Query Match             27.7%;  Score 54.5;  DB 1;  Length 198;
  Best Local Similarity   41.4%;  Pred. No. 19;
  Matches   12;  Conservative    3;  Mismatches    3;  Indels   11;  Gaps    1;

Qy      13 RAGG----------GCGGGGIEGPTLRQ 30
           ||||          | |||| :|||: :
Db     168 RAGGGVPSGMPGGMPGAGGGGGKGPTIEE 196


RESULT  14
```

```
OVO_DROME
ID   OVO_DROME      STANDARD;      PRT;  1028 AA.
AC   P51521; Q9XZU4;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   OVO PROTEIN (SHAVEN BABY PROTEIN).
GN   OVO OR SVB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=95021209; PubMed=7935398;
RA   Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT   "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT   melanogaster: relationship to genetic complexity.";
RL   Mol. Cell. Biol. 14:6809-6818(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=91293102; PubMed=1712294;
RA   Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT   "The ovo gene of Drosophila encodes a zinc finger protein required
RT   for female germ line development.";
RL   EMBO J. 10:2259-2266(1991).
CC   -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC       LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND
CC       ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC       BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC       PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11383; AAB60216.1; -.
DR   EMBL; X59772; CAB36921.1; ALT_SEQ.
DR   HSSP; P04002; 1WFA.
DR   FlyBase; FBgn0003028; ovo.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation.
FT   DOMAIN       62     66       POLY-ALA.
FT   DOMAIN       72     77       POLY-GLY.
FT   DOMAIN       80     85       POLY-GLY.
FT   DOMAIN       98    108       POLY-GLY.
FT   DOMAIN      144    152       POLY-HIS.
FT   DOMAIN      153    159       POLY-ASN.
FT   DOMAIN      336    339       POLY-GLN.
FT   DOMAIN      347    353       POLY-GLN.
FT   DOMAIN      357    361       POLY-GLN.
FT   DOMAIN      410    414       POLY-GLN.
FT   DOMAIN      418    422       POLY-GLN.
FT   DOMAIN      426    432       POLY-GLN.
FT   DOMAIN      445    453       POLY-GLN.
FT   DOMAIN      456    459       POLY-GLN.
FT   DOMAIN      466    474       POLY-GLN.
FT   DOMAIN      497    517       POLY-ALA.
FT   DOMAIN      524    529       POLY-SER.
FT   DOMAIN      549    558       POLY-ALA.
FT   DOMAIN      639    651       POLY-ALA.
FT   DOMAIN      717    725       POLY-ALA.
FT   DOMAIN      797    802       POLY-GLN.
FT   DOMAIN      820    823       POLY-GLN.
FT   DOMAIN      826    832       POLY-GLN.
FT   DOMAIN      874    992       ZINC FINGERS.
FT   ZN_FING     874    896       C2H2-TYPE.
FT   ZN_FING     902    924       C2H2-TYPE.
FT   ZN_FING     930    953       C2H2-TYPE.
FT   ZN_FING     969    992       C2H2-TYPE.
FT   CONFLICT    647    647       A -> R (IN REF. 2).
SQ   SEQUENCE   1028 AA;  110620 MW;  D7068BB2BC0F6F77 CRC64;


 Query Match             27.7%;  Score 54.5;  DB 1;  Length 1028;
 Best Local Similarity   57.9%;  Pred. No. 75;
 Matches   11;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    1;

Qy      11 AARAGGGC---GGGGIEGP 26
           |   ||||   ||||  ||
Db      71 AGSGGGGCTGNGGGGASGP 89


RESULT  15
YR21_TRSVR
ID   YR21_TRSVR     STANDARD;      PRT;   201 AA.
AC   P25245;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 20.2 KDA PROTEIN IN RNA2.
OS   Tomato ringspot virus (isolate raspberry) (Tomrsv).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC   Nepovirus.
OX   NCBI_TaxID=12281;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91311402; PubMed=1856689;
RA   Rott M.E., Tremaine J.H., Rochon D.M.;
RT   "Nucleotide sequence of tomato ringspot virus RNA-2.";
RL   J. Gen. Virol. 72:1505-1514(1991).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12477; BAA02044.1; -.
DR   PIR; JQ1094; JQ1094.
DR   HSSP; P04002; 1WFA.
KW   Hypothetical protein.
FT   DOMAIN       15     22       POLY-GLY.
FT   DOMAIN       61     66       POLY-GLY.
FT   DOMAIN      144    148       POLY-GLY.
SQ   SEQUENCE   201 AA;  20194 MW;  9038506E18D7B450 CRC64;


  Query Match             27.4%;  Score 54;  DB 1;  Length 201;
  Best Local Similarity   57.7%;  Pred. No. 21;
  Matches   15;  Conservative    1;  Mismatches    6;  Indels    4;  Gaps    1;

Qy      13 RAGGGCGGGGIE----GPTLRQWLAA 34
           |||||  |||| |    | || : | |
Db      13 RAGGGGGGGGKEVFKAGRTLLKVLKA 38


RESULT  16
Y870_METJA
```

```
ID   Y870_METJA     STANDARD;      PRT;   620 AA.
AC   Q58280;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0870.
GN   MJ0870.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: TO COENZYME F420 HYDROGENASE BETA SUBUNIT.
CC   -!- SIMILARITY: TO M.JANNASCHII MJ1349, MJ0725 AND MJ0551.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC       FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC       SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67531; AAB98876.1; -.
DR   HSSP; Q45560; 1BQX.
DR   TIGR; MJ0870; -.
DR   InterPro; IPR001450; 4FE4S_ferrdxin.
DR   InterPro; IPR000660; Nir_Sir.
DR   Pfam; PF00037; fer4; 3.
DR   Pfam; PF01077; NIR_SIR; 1.
DR   PRINTS; PR00397; SIROHAEM.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
DR   PROSITE; PS00365; NIR_SIR; 1.
KW   Hypothetical protein; Oxidoreductase; Heme; Iron-sulfur; 4Fe-4S;
KW   Complete proteome.
FT   METAL       428    428       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       434    434       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       468    468       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       472    472       IRON-SULFUR (4FE-4S) AND SIROHEME
FT                                 (BY SIMILARITY).
SQ   SEQUENCE   620 AA;  69793 MW;  9D71D2580D7D0BA8 CRC64;


  Query Match             27.4%;  Score 54;  DB 1;  Length 620;
  Best Local Similarity   43.5%;  Pred. No. 55;
  Matches   10;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 EGPTLRQWLAARAGGGCGGGGIE 24
           ||| :|  ||   || |  | ::
Db     418 EGPLVRATLACPGGGNCSSGLVD 440


RESULT  17
ORK1_DROME
ID   ORK1_DROME     STANDARD;      PRT;  1001 AA.
AC   Q94526;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   OPEN RECTIFIER POTASSIUM CHANNEL PROTEIN 1 (TWO PORE DOMAIN POTASSIUM
DE   CHANNEL ORK1).
GN   ORK1 OR CG1615.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Larva;
RX   MEDLINE=97075152; PubMed=8917578;
RA   Goldstein S.A.N., Price L.A., Rosenthal D.N., Pausch M.H.;
RT   "ORK1, a potassium-selective leak channel with two pore domains
RT   cloned from Drosophila melanogaster by expression in Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13256-13261(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: BACKGROUND POTASSIUM CHANNEL. RECTIFICATION IS DEPENDENT
CC       ON EXTERNAL POTASSIUM CONCENTRATION. ACTS AS AN OUTWARDLY
CC       RECTIFYING CHANNEL BUT AS EXTERNAL POTASSIUM LEVELS INCREASE, THIS
CC       IS REVERSED.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN ADULT, STRONGEST
CC       EXPRESSION IN MUSCLE, BRAIN AND OVARY. ALSO PRESENT AT LOW LEVELS
CC       IN LARVA AND EMBRYO.
CC   -!- MISCELLANEOUS: INHIBITED BY BARIUM.
CC   -!- SIMILARITY: BELONGS TO THE TWO PORE DOMAIN FAMILY OF POTASSIUM
CC       CHANNELS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U55321; AAC69250.1; -.
DR   EMBL; AE003484; AAF47972.1; -.
DR   FlyBase; FBgn0017561; Ork1.
DR   InterPro; IPR003280; 2poreK_channel.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR000099; TWIK_channel.
DR   Pfam; PF02034; TWIK_channel; 1.
DR   PRINTS; PR01333; 2POREKCHANEL.
KW   Ionic channel; Transmembrane; Ion transport; Potassium transport;
KW   Glycoprotein.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     27       POTENTIAL.
FT   DOMAIN       95    111       PORE-FORMING (POTENTIAL).
FT   TRANSMEM    120    140       POTENTIAL.
FT   DOMAIN      141    170       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    171    191       POTENTIAL.
FT   DOMAIN      208    224       PORE-FORMING (POTENTIAL).
FT   TRANSMEM    244    264       POTENTIAL.
FT   DOMAIN      265   1001       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     58     58       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1001 AA;  109289 MW;  09AE1A3669072E07 CRC64;
```

```
  Query Match             27.4%;  Score 54;  DB 1;  Length 1001;
  Best Local Similarity   52.2%;  Pred. No. 83;
  Matches   12;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGGIEGPTLRQ 30
           |  || |||  |||||   : :|
Db     761 QQQAAAAGGAAGGGGISRGSRKQ 783
```

```
RESULT  18
SCO2_HUMAN
ID   SCO2_HUMAN     STANDARD;      PRT;   266 AA.
AC   O43819; Q9UK87;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SCO2 PROTEIN HOMOLOG PRECURSOR.
GN   SCO2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Monocytes;
RA   Smink L.J., Burton J.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
RX   MEDLINE=20014747; PubMed=10545952;
RA   Papadopoulou L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
RA   Sadlock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,
RA   Van Coster R., Lyon G., Scalais E., Lebel R., Kaplan P., Shanske S.,
RA   De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;
RT   "Fatal infantile cardioencephalomyopathy with COX deficiency and
RT   mutations in SCO2, a COX assembly gene.";
RL   Nat. Genet. 23:333-337(1999).
CC   -!- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
CC       TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: UBUIQUITOUS.
CC   -!- DISEASE: DEFECTS IN SCO2 ARE THE CAUSE OF FATAL INFANTILE
CC       CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
CC       CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
CC       GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
CC       ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
CC       DEFICIENCIES.
CC   -!- SIMILARITY: BELONGS TO THE SCO1/2 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF177385; AAF05313.1; -.
DR   EMBL; AL021683; CAA16671.1; -.
DR   MIM; 604272; -.
DR   MIM; 604377; -.
DR   MIM; 220110; -.
DR   InterPro; IPR003782; SCO1_SenC.
DR   Pfam; PF02630; SCO1-SenC; 1.
KW   Mitochondrion; Transit peptide; Disease mutation.
FT   TRANSIT       1     41       MITOCHONDRION (POTENTIAL).
FT   CHAIN        42    266       SCO2 PROTEIN HOMOLOG.
FT   VARIANT     140    140       E -> K (IN FIC).
FT                                /FTId=VAR_008874.
FT   VARIANT     225    225       S -> F (IN FIC).
FT                                /FTId=VAR_008875.
FT   CONFLICT     20     20       R -> P (IN REF. 1).
SQ   SEQUENCE   266 AA;  29810 MW;  BC2F40E057329BF3 CRC64;
```

```
  Query Match             27.2%;  Score 53.5;  DB 1;  Length 266;
  Best Local Similarity   35.4%;  Pred. No. 30;
  Matches   17;  Conservative    2;  Mismatches   12;  Indels   17;  Gaps    2;

Qy       6 LRQWLAARAGGG--CGGGGIEGPTLR--------------QWLAARA 36
           || || :| |     | |  :|| ||                 ||| ||
Db      33 LRSWLLSRQGPAETGGQGQPQGPGLRTRLLITGLFGAGLGGAWLALRA 80
```

```
RESULT  19
SYK_AERPE
ID   SYK_AERPE      STANDARD;      PRT;   562 AA.
AC   Q9YFT9;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LYSYL-TRNA SYNTHETASE (EC 6.1.1.6) (LYSINE--TRNA LIGASE) (LYSRS).
GN   LYSS OR APE0161.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + L-LYSINE + TRNA(LYS) = AMP +
CC       PYROPHOSPHATE + L-LYSYL-TRNA(LYS).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000058; BAA79072.1; -.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002904; tRNA-synt_lys_1.
DR   Pfam; PF01921; tRNA-synt_1f; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         50     58       "HIGH" REGION.
FT   SITE        305    309       "KMSKS" REGION.
SQ   SEQUENCE   562 AA;  65114 MW;  753664E2937FBF27 CRC64;


  Query Match             27.2%;  Score 53.5;  DB 1;  Length 562;
  Best Local Similarity   39.3%;  Pred. No. 58;
  Matches   11;  Conservative    4;  Mismatches   10;  Indels    3;  Gaps    1;

Qy       8 QWLAARAGG---GCGGGGIEGPTLRQWL 32
           :|:: ||||        |  | | |:||
Db     293 EWVSLRAGGREADMSSSGFTGITPREWL 320


RESULT  20
GATA_MYCLE
ID   GATA_MYCLE     STANDARD;      PRT;   497 AA.
AC   O33105;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT A (EC 6.3.5.-) (GLU-ADT
DE   SUBUNIT A).
GN   GATA OR ML1702 OR MLCB637.13.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC       GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
CC       TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC       REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC       AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC       + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC   -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z99263; CAB16428.1; -.
DR   EMBL; AL583923; CAC30655.1; -.
DR   Leproma; ML1702; -.
DR   InterPro; IPR000120; Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Protein biosynthesis; Ligase; Complete proteome.
SQ   SEQUENCE   497 AA;  51536 MW;  D3723D871518BDC7 CRC64;


  Query Match             26.9%;  Score 53;  DB 1;  Length 497;
  Best Local Similarity   52.6%;  Pred. No. 59;
  Matches   10;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       3 GPTLRQWLAARAGGGCGGG 21
           |||   |   |  || |||
Db     145 GPTRNPWNVDRVPGGSGGG 163


RESULT  21
E2BE_RAT
ID   E2BE_RAT       STANDARD;      PRT;   716 AA.
AC   Q64350;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRANSLATION INITIATION FACTOR EIF-2B EPSILON SUBUNIT (EIF-2B GDP-GTP
DE   EXCHANGE FACTOR).
GN   EIF2B5 OR EIF2BE.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=96305355; PubMed=8688467;
RA   Flowers K.M., Mellor H., Matts R.L., Kimball S.R., Jefferson L.S.;
RT   "Cloning and characterization of complementary and genomic DNAs
RT   encoding the epsilon-subunit of rat translation initiation
RT   factor-2B.";
RL   Biochim. Biophys. Acta 1307:318-324(1996).
CC   -!- FUNCTION: CATALYZES THE EXCHANGE OF EUKARYOTIC INITIATION FACTOR
CC       2-BOUND GDP FOR GTP.
CC   -!- SUBUNIT: COMPLEX OF FIVE DIFFERENT SUBUNITS; ALPHA, BETA, GAMMA,
CC       DELTA AND EPSILON.
CC   -!- SIMILARITY: BELONGS TO THE EIF-2B GAMMA/EPSILON SUBUNITS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U19516; AAB17690.1; -.
DR   EMBL; U19511; AAB17691.1; -.
DR   InterPro; IPR001451; Hexapep_transf.
DR   InterPro; IPR003307; eIF5C.
DR   Pfam; PF00132; hexapep; 3.
DR   Pfam; PF02020; W2; 1.
DR   SMART; SM00515; eIF5C; 1.
KW   Amino-acid biosynthesis; Translation regulation.
FT   DOMAIN       19     26       POLY-GLY.
FT   DOMAIN       34     37       POLY-PRO.
SQ   SEQUENCE   716 AA;  80240 MW;  C6E4BFCE060AF6F1 CRC64;
```

```
  Query Match             26.9%;  Score 53;  DB 1;  Length 716;
  Best Local Similarity   43.3%;  Pred. No. 80;
  Matches   13;  Conservative    3;  Mismatches    8;  Indels    6;  Gaps    1;

Qy      11 AARAGGGCGGGGIEG------PTLRQWLAA 34
           | : ||| |||| :|      | |:  | |
Db      15 ANKRGGGSGGGGTQGAEEEPPPPLQAVLVA 44


RESULT  22
SSB_RHOSH
ID   SSB_RHOSH      STANDARD;      PRT;   174 AA.
AC   Q9ZAQ8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SINGLE-STRAND BINDING PROTEIN (SSB) (HELIX-DESTABILIZING PROTEIN).
GN   SSB.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RA   Zeilstra-Ryalls J.H., Kaplan S.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
CC       CHROMOSOME. IT IS ALSO INVOLVED IN DNA RECOMBINATION AND REPAIR
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SSB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82280; AAD00529.1; -.
DR   HSSP; P02339; 1KAW.
DR   InterPro; IPR000424; SSB.
DR   Pfam; PF00436; SSB; 1.
DR   PROSITE; PS00735; SSB_1; FALSE_NEG.
DR   PROSITE; PS00736; SSB_2; FALSE_NEG.
KW   DNA-binding; DNA repair; DNA replication.
SQ   SEQUENCE   174 AA;  18496 MW;  DBF5BC8D034D532D CRC64;


  Query Match             26.6%;  Score 52.5;  DB 1;  Length 174;
  Best Local Similarity   66.7%;  Pred. No. 27;
  Matches   12;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    1;

Qy      12 ARAGGGCGGGGIE---GP 26
           | |||| |||| |   ||
Db     122 AGAGGGMGGGGYEDRGGP 139


RESULT  23
SPIN_CBEPV
ID   SPIN_CBEPV     STANDARD;      PRT;   341 AA.
AC   P23061;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SPINDOLIN PRECURSOR (SPHEROIDIN).
GN   P50 OR SPH.
OS   Choristoneura biennis entomopoxvirus (CbEPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC   Entomopoxvirus B.
OX   NCBI_TaxID=10288;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 21-50.
RX   MEDLINE=90223988; PubMed=2327073;
RA   Yuen L., Dionne J., Arif B., Richardson C.;
RT   "Identification and sequencing of the spheroidin gene of
RT   Choristoneura biennis entomopoxvirus.";
RL   Virology 175:427-433(1990).
RN   [2]
RP   REVISION TO FUNCTION.
RX   MEDLINE=93389435; PubMed=8376960;
RA   Dall D., Sriskantha A., Vera A., Lai-Fook J., Symonds T.;
RT   "A gene encoding a highly expressed spindle body protein of Heliothis
RT   armigera entomopoxvirus.";
RL   J. Gen. Virol. 74:1811-1818(1993).
CC   -!- FUNCTION: THIS PROTEIN IS A SPINDLE BODY PROTEIN.
CC   -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: WITH HAEPV SPINDOLIN AND ACMNPV SPINDOLIN-LIKE
CC       PROTEIN.
CC   -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A SPHEROIDIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34140; AAA42887.1; -.
DR   PIR; A34743; PYVZCB.
KW   Signal; Late protein.
FT   SIGNAL        1     20
FT   CHAIN        21    341       SPINDOLIN.
SQ   SEQUENCE   341 AA;  38709 MW;  E84EF9BCD901E72F CRC64;


  Query Match             26.6%;  Score 52.5;  DB 1;  Length 341;
  Best Local Similarity   44.8%;  Pred. No. 48;
  Matches   13;  Conservative    2;  Mismatches   11;  Indels    3;  Gaps    1;

Qy       4 PTLRQWLAARAGGG---CGGGGIEGPTLR 29
           |  ||   : |||     |||||: |  |
Db      27 PIARQRRCSAAGGNWYPVGGGGIQDPMCR 55


RESULT  24
CYB_MICIK
ID   CYB_MICIK      STANDARD;      PRT;   370 AA.
AC   Q9MLK2;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CYTOCHROME B.
GN   MTCYB OR COB OR CYTB.
OS   Micropechis ikaheka.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Notechinae; Micropechis.
OX   NCBI_TaxID=66188;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20229584; PubMed=10764543;
RA   Slowinski J.B., Keogh J.S.;
RT   "Phylogenetic relationships of elapid snakes based on cytochrome b
RT   mtDNA sequences.";
RL   Mol. Phylogenet. Evol. 15:157-164(2000).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN (BY SIMILARITY).
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
```

```
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF217826; AAF37245.1; -.
DR   InterPro; IPR000179; Cyt_b_b6.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; FALSE_NEG.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        75     75       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        89     89       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       174    174       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       188    188       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   370 AA;  42083 MW;  CCDE45269CAB2B9D CRC64;


  Query Match             26.6%;  Score 52.5;  DB 1;  Length 370;
  Best Local Similarity   41.9%;  Pred. No. 51;
  Matches   13;  Conservative    2;  Mismatches    9;  Indels    7;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPTLRQWLA 33
           ||||  ||         ||  |  ||| :: |
Db     149 GPTLTTWL-------WGGFSINDPTLTRFFA 172


RESULT  25
RU17_DROME
ID   RU17_DROME     STANDARD;      PRT;   448 AA.
AC   P17133; Q9VM56;
DT   01-AUG-1990 (Rel. 15, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA (U1 SNRNP 70 KDA) (SNRNP70).
GN   SNRNP70K OR SNRNP27D OR CG8749.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258833; PubMed=1692955;
RA   Mancebo R., Lo P.C.H., Mount S.M.;
RT   "Structure and expression of the Drosophila melanogaster gene for the
RT   U1 small nuclear ribonucleoprotein particle 70K protein.";
RL   Mol. Cell. Biol. 10:2492-2502(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE STEM
CC       LOOP I REGION OF U1-SNRNA.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31162; AAA28859.1; -.
DR   EMBL; AE003615; AAF52471.1; -.
DR   PIR; A36311; A36311.
DR   HSSP; P09651; 1UP1.
DR   FlyBase; FBgn0016978; snRNP70K.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing.
FT   DOMAIN      102    180       RNA-BINDING (RRM).
FT   DOMAIN      254    350       ARG/GLU-RICH (MIXED CHARGE).
FT   CONFLICT    278    278       N -> S (IN REF. 1).
SQ   SEQUENCE   448 AA;  52900 MW;  0DDFB5A39CA72AEB CRC64;


  Query Match             26.6%;  Score 52.5;  DB 1;  Length 448;
  Best Local Similarity   52.4%;  Pred. No. 61;
  Matches   11;  Conservative    2;  Mismatches    5;  Indels    3;  Gaps    1;

Qy       5 TLRQWLAARAGGCGG---GG 22
           |:: ||  | ||| ||   ||
Db     182 TVKGWLPRRLGGGLGGTRRGG 202


RESULT  26
PAC4_HUMAN
ID   PAC4_HUMAN     STANDARD;      PRT;   969 AA.
AC   P29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
AC   Q9UEG7; Q9Y4G9; Q9Y4H0; Q9Y4H1;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PAIRED BASIC AMINO ACID CLEAVING ENZYME 4 PRECURSOR (EC 3.4.21.-)
DE   (SUBTILISIN/KEXIN-LIKE PROTEASE PACE4) (SUBTILISIN-LIKE PROPROTEIN
```

```
DE   CONVERTASE 4) (SPC4).
GN   PACE4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
RC   TISSUE=Hepatoma, and Kidney;
RX   MEDLINE=92075167; PubMed=1741956;
RA   Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
RA   Barr P.J.;
RT   "Identification of a second human subtilisin-like protease gene in
RT   the fes/fps region of chromosome 15.";
RL   DNA Cell Biol. 10:757-769(1991).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
RC   TISSUE=Placenta;
RX   MEDLINE=94235049; PubMed=8179631;
RA   Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
RA   Matsuda Y.;
RT   "Identification of novel cDNAs encoding human kexin-like protease,
RT   PACE4 isoforms.";
RL   Biochem. Biophys. Res. Commun. 200:943-950(1994).
RN   [3]
RP   ERRATUM.
RX   MEDLINE=95071480; PubMed=7980617;
RA   Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
RA   Matsuda Y.;
RT   "Identification of novel cDNAs encoding human kexin-like protease,
RT   PACE4 isoforms.";
RL   Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
RC   TISSUE=Placenta;
RA   Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
RA   Matsuda Y.;
RT   "Identification of a novel PACE4 isoform, PACE4E.";
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
RC   TISSUE=Cerebellum;
RX   MEDLINE=97335942; PubMed=9192737;
RA   Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
RA   Akamatsu T., Nagamune H., Matsuda Y.;
RT   "A novel human PACE4 isoform, PACE4E is an active processing protease
RT   containing a hydrophobic cluster at the carboxy terminus.";
RL   J. Biochem. 121:941-948(1997).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
RX   MEDLINE=98021085; PubMed=9378725;
RA   Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
RA   Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
RT   "Genomic organization and alternative splicing of human PACE4 (SPC4),
RT   kexin-like processing endoprotease.";
RL   J. Biochem. 122:438-452(1997).
RN   [7]
RP   ALTERNATIVE SPLICING (ISOFORM PACE4CS).
RX   MEDLINE=97064242; PubMed=8906861;
RA   Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
RT   "Functional analysis of human PACE4-A and PACE4-C isoforms:
RT   identification of a new PACE4-CS isoform.";
RL   FEBS Lett. 396:31-36(1996).
RN   [8]
RP   CHARACTERIZATION.
RX   MEDLINE=99233559; PubMed=10215603;
RA   Sucic J.F., Moehring J.M., Inocencio N.M., Luchini J.W.,
RA   Moehring T.J.;
RT   "Endoprotease PACE4 is Ca2+-dependent and temperature-sensitive and
RT   can partly rescue the phenotype of a furin-deficient cell strain.";
RL   Biochem. J. 339:639-647(1999).
RN   [9]
RP   PROCESSING.
RX   MEDLINE=98408849; PubMed=9738469;
RA   Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
RA   Tsuji A., Matsuda Y.;
RT   "Biosynthetic processing and quaternary interactions of proprotein
RT   convertase SPC4 (PACE4).";
RL   FEBS Lett. 434:155-159(1998).
CC   -!- FUNCTION: LIKELY TO REPRESENT AN ENDOPROTEASE ACTIVITY WITHIN THE
CC       CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
CC       DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
CC       AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
CC   -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC       PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS,
CC       WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC   -!- COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.
CC   -!- SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
CC       RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
CC       WHEREAS MATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
CC       PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: PACE4A-I AND PACE4-AII ARE SECRETED. PACE4C
CC       AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
CC       IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
CC       INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
CC       TERMINUS. PACE4B MIGHT BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS; PACE4A-I/PACE4 (SHOWN HERE),
CC       PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
CC       PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
CC       C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
CC   -!- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
CC       RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
CC       PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
CC       COMPARATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
CC       EXPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
CC       KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE
CC       EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM,
CC       PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN
CC       CEREBELLUM.
CC   -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC       ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC       RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M80482; AAA59998.1; -.
DR   EMBL; AB001914; BAA21620.1; -.
DR   EMBL; AB001898; BAA21620.1; JOINED.
DR   EMBL; AB001900; BAA21620.1; JOINED.
DR   EMBL; AB001901; BAA21620.1; JOINED.
DR   EMBL; AB001902; BAA21620.1; JOINED.
DR   EMBL; AB001903; BAA21620.1; JOINED.
DR   EMBL; AB001904; BAA21620.1; JOINED.
DR   EMBL; AB001905; BAA21620.1; JOINED.
DR   EMBL; AB001914; BAA21621.1; -.
DR   EMBL; AB001898; BAA21621.1; JOINED.
DR   EMBL; AB001900; BAA21621.1; JOINED.
DR   EMBL; AB001901; BAA21621.1; JOINED.
DR   EMBL; AB001902; BAA21621.1; JOINED.
DR   EMBL; AB001903; BAA21621.1; JOINED.
DR   EMBL; AB001904; BAA21621.1; JOINED.
DR   EMBL; AB001905; BAA21621.1; JOINED.
DR   EMBL; AB001906; BAA21621.1; JOINED.
DR   EMBL; AB001907; BAA21621.1; JOINED.
DR   EMBL; AB001908; BAA21621.1; JOINED.
DR   EMBL; AB001909; BAA21621.1; JOINED.
DR   EMBL; AB001914; BAA21622.1; -.
```

```
DR   EMBL; AB001901; BAA21622.1; JOINED.
DR   EMBL; AB001902; BAA21622.1; JOINED.
DR   EMBL; AB001903; BAA21622.1; JOINED.
DR   EMBL; AB001904; BAA21622.1; JOINED.
DR   EMBL; AB001905; BAA21622.1; JOINED.
DR   EMBL; AB001906; BAA21622.1; JOINED.
DR   EMBL; AB001907; BAA21622.1; JOINED.
DR   EMBL; AB001908; BAA21622.1; JOINED.
DR   EMBL; AB001914; BAA21623.1; -.
DR   EMBL; AB001898; BAA21623.1; JOINED.
DR   EMBL; AB001900; BAA21623.1; JOINED.
DR   EMBL; AB001901; BAA21623.1; JOINED.
DR   EMBL; AB001902; BAA21623.1; JOINED.
DR   EMBL; AB001903; BAA21623.1; JOINED.
DR   EMBL; AB001904; BAA21623.1; JOINED.
DR   EMBL; AB001905; BAA21623.1; JOINED.
DR   EMBL; AB001906; BAA21623.1; JOINED.
DR   EMBL; AB001907; BAA21623.1; JOINED.
DR   EMBL; AB001908; BAA21623.1; JOINED.
DR   EMBL; AB001909; BAA21623.1; JOINED.
DR   EMBL; AB001914; BAA21624.1; -.
DR   EMBL; AB001898; BAA21624.1; JOINED.
DR   EMBL; AB001900; BAA21624.1; JOINED.
DR   EMBL; AB001901; BAA21624.1; JOINED.
DR   EMBL; AB001902; BAA21624.1; JOINED.
DR   EMBL; AB001903; BAA21624.1; JOINED.
DR   EMBL; AB001904; BAA21624.1; JOINED.
DR   EMBL; AB001905; BAA21624.1; JOINED.
DR   EMBL; AB001906; BAA21624.1; JOINED.
DR   EMBL; AB001907; BAA21624.1; JOINED.
DR   EMBL; AB001908; BAA21624.1; JOINED.
DR   EMBL; AB001910; BAA21624.1; JOINED.
DR   EMBL; AB001911; BAA21624.1; JOINED.
DR   EMBL; AB001912; BAA21624.1; JOINED.
DR   EMBL; AB001913; BAA21624.1; JOINED.
DR   EMBL; AB001914; BAA21625.1; -.
DR   EMBL; AB001898; BAA21625.1; JOINED.
DR   EMBL; AB001900; BAA21625.1; JOINED.
DR   EMBL; AB001901; BAA21625.1; JOINED.
DR   EMBL; AB001902; BAA21625.1; JOINED.
DR   EMBL; AB001903; BAA21625.1; JOINED.
DR   EMBL; AB001904; BAA21625.1; JOINED.
DR   EMBL; AB001905; BAA21625.1; JOINED.
DR   EMBL; AB001906; BAA21625.1; JOINED.
DR   EMBL; AB001907; BAA21625.1; JOINED.
DR   EMBL; AB001908; BAA21625.1; JOINED.

  Query Match             26.6%;  Score 52.5;  DB 1;  Length 969;
  Best Local Similarity   44.8%;  Pred. No. 1.2e+02;
  Matches   13;  Conservative    1;  Mismatches    8;  Indels    7;  Gaps    1;

Qy      11 AARAGGGCGGGGIEGPTLR-------QWL 32
           || |||  | ||  ||  |       :||
Db      24 AAGAGGAGGAGGAGGPGFRPLAPRPWRWL 52


RESULT  27
SIX3_MOUSE
ID   SIX3_MOUSE     STANDARD;      PRT;   333 AA.
AC   Q62233; P70176; P70177;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HOMEOBOX PROTEIN SIX3 (SINE OCULIS HOMEOBOX HOMOLOG 3).
GN   SIX3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Embryonic brain;
RX   MEDLINE=96125147; PubMed=8575305;
RA   Oliver G., Mailhos A., Wehr R., Copeland N.G., Jenkins N.A.,
RA   Gruss P.;
RT   "Six3, a murine homologue of the sine oculis gene, demarcates the
RT   most anterior border of the developing neural plate and is expressed
RT   during eye development.";
RL   Development 121:4045-4055(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=96409319; PubMed=8814301;
RA   Kawakami K., Ohto H., Takizawa T., Saito T.;
RT   "Identification and expression of Six family genes in mouse retina.";
RL   FEBS Lett. 393:259-263(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN VISUAL SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIX3A AND SIX3B (SHOWN HERE);
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT E6.5 OF EMBRYO DEVELOPMENT
CC       AROUND THE ANTERIOR BORDER. AT E8.5, EXPRESSION IS FOUND OVER THE
CC       ANTERIOR NEURAL PLATE. AT E9.5, IN THE DIENCEPHALIC PART OF THE
CC       VENTRAL FOREBRAIN, OPTIC VESICLES, OLFACTORY PLACODES AND RATHKE'S
CC       POUCH. IN LATER STAGES, PRESENT IN HYPOTHALAMUS, EYES AND
CC       PITUITARY.
CC   -!- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEODOMAIN
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X90871; CAA62379.1; ALT_INIT.
DR   EMBL; D83144; BAA11822.1; -.
DR   EMBL; D83145; BAA11823.1; -.
DR   MGD; MGI:102764; Six3.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; FALSE_NEG.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW   Alternative splicing.
FT   DOMAIN       33     70       GLY-RICH.
FT   DNA_BIND     207    266      HOMEOBOX.
FT   DOMAIN       264    267      POLY-ALA.
FT   VARSPLIC     271    286      LQHQAIGPSGMRSLAE -> SVAGTAARPPQA (IN
FT                                ISOFORM SIX3A).
FT   VARSPLIC     287    333      MISSING (IN ISOFORM SIX3A).
FT   CONFLICT      44     44      G -> GG (IN REF. 1).
FT   CONFLICT     118    119      VA -> WP (IN REF. 1).
FT   CONFLICT     278    333      PSGMRSLAEPGCPTHGSAESPSTAASPTTSVSSLTERADTG
FT                                TSILSVTSSDSECDV -> ERDALPGRARLPHARLSRVTVH
FT                                GGQPDHQCVQPDGARGHRHFDPLGNLQRLGM (IN REF.
FT                                1).
SQ   SEQUENCE   333 AA;  35592 MW;  1AD7D3C4388043B9 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 333;
  Best Local Similarity   62.5%;  Pred. No. 53;
  Matches   10;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      11 AARAGGGCGGGGIEGP 26
           |  |||| ||||   |
Db      59 AGGAGGGSGGGGSRAP 74


RESULT  28
FXD3_CHICK
ID   FXD3_CHICK     STANDARD;      PRT;   394 AA.
AC   P79772;
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FORKHEAD BOX PROTEIN D3 (HNF3/FH TRANSCRIPTION FACTOR GENESIS) (WINGED
DE   HELIX PROTEIN CWH-3).
GN   FOXD3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=97141794; PubMed=8988052;
RA   Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
RT   "Aberrant cell growth induced by avian winged helix proteins.";
RL   Cancer Res. 57:123-129(1997).
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U37274; AAC60066.1; -.
DR   InterPro; IPR001766; Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation.
FT   DOMAIN       67     70       POLY-ALA.
FT   DOMAIN       80     91       POLY-GLY.
FT   DOMAIN      100    106       POLY-ALA.
FT   DNA_BIND    117    211       FORK-HEAD.
SQ   SEQUENCE   394 AA;  40995 MW;  324A4B36B9E31899 CRC64;
```

```
  Query Match             26.4%;  Score 52;  DB 1;  Length 394;
  Best Local Similarity   76.9%;  Pred. No. 61;
  Matches   10;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      13 RAGGGCGGGGIEG 25
           | ||| |||| ||
Db      82 RGGGGGGGGGEEG 94
```

```
RESULT  29
HKLB_LYCES
ID   HKLB_LYCES     STANDARD;      PRT;   426 AA.
AC   O22300;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HOMEOBOX PROTEIN KNOTTED-1 LIKE LET12.
GN   LET12.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VFNT CHERRY;
RX   MEDLINE=98145476; PubMed=9484482;
RA   Janssen B.J., Williams A., Chen J.J., Mathern J., Hake S., Sinha N.;
RT   "Isolation and characterization of two knotted-like homeobox genes
RT   from tomato.";
RL   Plant Mol. Biol. 36:417-425(1998).
CC   -!- FUNCTION: MAY HAVE A ROLE TO PLAY IN FORMATIVE EVENTS IN OVULE AND
CC       EMBRYO MORPHOGENESIS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN THE MATURE PLANT.
CC   -!- SIMILARITY: BELONGS TO THE TALE/KNOX FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF000142; AAC49918.1; -.
DR   InterPro; IPR001356; Homeobox.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
FT   DOMAIN       15     24       POLY-GLN.
FT   DOMAIN       69     76       POLY-ALA.
FT   DOMAIN      140    152       POLY-ASN.
FT   DOMAIN      283    287       POLY-ASP.
FT   DOMAIN      325    348       ELK DOMAIN.
FT   DNA_BIND    349    411       HOMEOBOX (TALE-TYPE).
SQ   SEQUENCE   426 AA;  47581 MW;  5B52B9E0A34A86BC CRC64;
```

```
  Query Match             26.4%;  Score 52;  DB 1;  Length 426;
  Best Local Similarity   64.7%;  Pred. No. 66;
  Matches   11;  Conservative    1;  Mismatches    3;  Indels    2;  Gaps    1;

Qy       8 QWLA--ARAGGGCGGGG 22
           |||:  | ||||  |||
Db      96 QWLSPTAAAGGGSNGGG 112
```

```
RESULT  30
OC3N_HUMAN
ID   OC3N_HUMAN     STANDARD;      PRT;   443 AA.
AC   P20265; Q14960;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NERVOUS-SYSTEM SPECIFIC OCTAMER-BINDING TRANSCRIPTION FACTOR N-OCT 3
DE   (BRAIN-SPECIFIC HOMEOBOX/POU DOMAIN PROTEIN 2) (BRN-2 PROTEIN)
DE   [CONTAINS: N-OCT 5A; N-OCT 5B].
GN   POU3F2 OR BRN2 OR OTF7 OR OCT7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93181199; PubMed=8441633;
RA   Schreiber E., Tobler A., Malipiero U., Schaffner W., Fontana A.;
RT   "cDNA cloning of human N-Oct3, a nervous-system specific POU domain
RT   transcription factor binding to the octamer DNA motif.";
RL   Nucleic Acids Res. 21:253-258(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95380176; PubMed=7651733;
RA   Angus J., Thomson F., Murphy K., Baker E., Sutherland G.R.,
RA   Parsons P.G., Sturm R.A.;
RT   "The brn-2 gene regulates the melanocytic phenotype and tumorigenic
RT   potential of human melanoma cells.";
```

```
RL   Oncogene 11:691-700(1995).
RN   [3]
RP   SEQUENCE OF 280-404 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=89295573; PubMed=2739723;
RA   He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
RA   Rosenfeld M.G.;
RT   "Expression of a large family of POU-domain regulatory genes in
RT   mammalian brain development.";
RL   Nature 340:35-42(1989).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
CC       RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
CC       ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
CC       OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
CC       THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
CC       PROMOTERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; N-OCT 3 (SHOWN HERE), N-OCT 5A
CC       AND N-OCT 5B; ARE PRODUCED BY ALTERNATIVE INITIATION.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
CC       CELL LINEAGE.
CC   -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
CC       TO CLASS-3 POU.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11933; CAA77990.1; -.
DR   EMBL; L37868; AAB59611.1; -.
DR   PIR; S05043; S05043.
DR   PIR; S29334; S29334.
DR   HSSP; P14859; 1POU.
DR   TRANSFAC; T00630; -.
DR   MIM; 600494; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000327; POU.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00157; pou; 1.
DR   PRINTS; PR00028; POUDOMAIN.
DR   ProDom; PD000583; POU; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00352; POU; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS00035; POU_1; 1.
DR   PROSITE; PS00465; POU_2; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
KW   Activator; Alternative initiation.
FT   CHAIN         1    443       N-OCT 3.
FT   CHAIN       181    443       N-OCT 5A.
FT   CHAIN       200    443       N-OCT 5B.
FT   INIT_MET    181    181       FOR N-OCT 5A.
FT   INIT_MET    200    200       FOR N-OCT 5B.
FT   DOMAIN       68     90       POLY-GLY.
FT   DOMAIN      125    149       POLY-GLN.
FT   DOMAIN      266    336       POU.
FT   DNA_BIND    354    413       HOMEOBOX.
FT   CONFLICT     26     26       A -> G (IN REF. 2).
SQ   SEQUENCE   443 AA;  46921 MW;  2CAC852328334A66 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 443;
  Best Local Similarity   60.0%;  Pred. No. 68;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGG 22
           ||: | : || ||||
Db      60 QWITALSHGGGGGGG 74
```

```
RESULT  31
OC3N_MOUSE
ID   OC3N_MOUSE     STANDARD;      PRT;   445 AA.
AC   P31360;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NERVOUS-SYSTEM SPECIFIC OCTAMER-BINDING TRANSCRIPTION FACTOR N-OCT 3
DE   (BRAIN-SPECIFIC HOMEOBOX/POU DOMAIN PROTEIN 2) (BRN-2 PROTEIN).
GN   POU3F2 OR OTF7 OR BRN2 OR BRN-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92228768; PubMed=1565620;
RA   Hara Y., Rovescalli C., Kim Y., Nirenberg M.;
RT   "Structure and evolution of four POU domain genes expressed in mouse
RT   brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
CC       RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
CC       ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
CC       OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
CC       THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
CC       PROMOTERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
CC       CELL LINEAGE.
CC   -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
CC       TO CLASS-3 POU.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88300; AAA39961.1; -.
DR   PIR; S31224; S31224.
DR   HSSP; P14859; 1POU.
DR   MGD; MGI:101895; Pou3f2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000327; POU.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00157; pou; 1.
DR   PRINTS; PR00028; POUDOMAIN.
DR   ProDom; PD000583; POU; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00352; POU; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PROSITE; PS00035; POU_1; 1.
DR   PROSITE; PS00465; POU_2; 1.
KW   DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
KW   Activator.
FT   DOMAIN       68     90       POLY-GLY.
FT   DOMAIN      125    149       POLY-GLN.
FT   DOMAIN      268    338       POU.
FT   DNA_BIND    356    415       HOMEOBOX.
SQ   SEQUENCE   445 AA;  47149 MW;  1A47F10950EECE8A CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 445;
  Best Local Similarity   60.0%;  Pred. No. 68;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       8 QWLAARAGGGCGGGG 22
           ||: | : || ||||
Db      60 QWITALSHGGGGGGG 74


RESULT  32
SRF_XENLA
ID   SRF_XENLA      STANDARD;      PRT;   448 AA.
AC   P23790;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   SERUM RESPONSE FACTOR (SRF).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91184140; PubMed=2009862;
RA   Mohun T.J., Chambers A.E., Towers N., Taylor M.V.;
RT   "Expression of genes encoding the transcription factor SRF during
RT   early development of Xenopus laevis: identification of a CArG
RT   box-binding activity as SRF.";
RL   EMBO J. 10:933-940(1991).
CC   -!- FUNCTION: SRF IS A TRANSCRIPTION FACTOR THAT BINDS TO THE SERUM
CC       RESPONSE ELEMENT (SRE), A SHORT SEQUENCE OF DYAD SYMMETRY LOCATED
CC       300 BP TO THE 5' OF THE SITE OF TRANSCRIPTION INITIATION OF SOME
CC       GENES.
CC   -!- SUBUNIT: BINDS DNA AS A MULTIMER, PROBABLY A DIMER.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- PTM: PHOSPHORYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC       FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56451; CAA39832.1; -.
DR   PIR; S15018; S15018.
DR   HSSP; P11831; 1SRS.
DR   TRANSFAC; T00763; -.
DR   InterPro; IPR002100; MADS-box.
DR   Pfam; PF00319; SRF-TF; 1.
DR   PRINTS; PR00404; MADSDOMAIN.
DR   SMART; SM00432; MADS; 1.
DR   PROSITE; PS00350; MADS_BOX_1; 1.
DR   PROSITE; PS50066; MADS_BOX_2; 1.
KW   Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW   Phosphorylation.
FT   DOMAIN       98    152       MADS.
SQ   SEQUENCE   448 AA;  46115 MW;  B3CDCA7E0D97C23B CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 448;
  Best Local Similarity   64.7%;  Pred. No. 68;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      10 LAARAGGGCGGGGIEGP 26
           || ||| | |  || ||
Db      18 LARRAGNGAGCPGIRGP 34


RESULT  33
ERR1_MOUSE
ID   ERR1_MOUSE     STANDARD;      PRT;   462 AA.
AC   O08580;
DT   15-JUL-1999 (Rel. 38, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   STEROID HORMONE RECEPTOR ERR1 (ESTROGEN-RELATED RECEPTOR, ALPHA)
DE   (ERR-ALPHA) (ESTROGEN RECEPTOR-LIKE 1) (FRAGMENT).
GN   ESRRA OR NR3B1 OR ERR1 OR ESTRRA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=97415618; PubMed=9271417;
RA   Sladek R., Bader J.-A., Giguere V.;
RT   "The orphan nuclear receptor estrogen-related receptor alpha is a
RT   transcriptional regulator of the human medium-chain acyl coenzyme A
RT   dehydrogenase gene.";
RL   Mol. Cell. Biol. 17:5400-5409(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Kidney;
RX   MEDLINE=98121983; PubMed=9460651;
RA   Shigeta H., Zuo W., Yang N., DiAugustine R., Teng C.T.;
RT   "The mouse estrogen receptor-related orphan receptor alpha 1:
RT   molecular cloning and estrogen responsiveness.";
RL   J. Mol. Endocrinol. 19:299-309(1997).
CC   -!- FUNCTION: BINDS TO AN ERR-ALPHA RESPONSE ELEMENT (ERRE) CONTAINING
CC       A SINGLE CONSENSUS HALF-SITE, 5'-TNAAGGTCA-3'. CAN BIND TO THE
CC       MEDIUM-CHAIN ACYL COENZYME A DEHYDROGENASE (MCAD) RESPONSE ELEMENT
CC       NRRE-1 AND MAY ACT AS AN IMPORTANT REGULATOR OF MCAD PROMOTER. MAY
CC       FUNCTION AS A MODULATOR OF THE ESTROGEN SIGNALING PATHWAY IN THE
CC       UTERUS.
CC   -!- SUBUNIT: BINDS DNA AS A MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: MOST HIGHLY EXPRESSED IN KIDNEY, HEART, AND
CC       BROWN ADIPOCYTES. ALSO FOUND IN UTERUS, CERVIX AND VAGINA.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN AN ORGAN SPECIFIC MANNER THROUGH
CC       MID-TO LATE EMBRYONIC DEVELOPMENT WITH PERSISTENT HIGH-LEVEL
CC       EXPRESSION IN BROWN ADIPOSE TISSUE AND INTESTINAL MUCOSA.
CC   -!- INDUCTION: ACTIVATED BY DIETHYLSTILBESTROL (DES) AND ESTRADIOL IN
CC       THE UTERUS.
CC   -!- PTM: PHOSPHORYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U85259; AAB51250.1; ALT_INIT.
DR   MGD; MGI:1346831; Esrra.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Strdhormone_rcptor.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   PRINTS; PR00350; VITAMINDR.
DR   PRINTS; PR00398; STRDHORMONER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation.
FT   NON_TER       1      1
FT   DNA_BIND    119    184       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     119    139       C4-TYPE.
FT   ZN_FING     155    179       C4-TYPE.
```

```
SQ   SEQUENCE   462 AA;  49280 MW;  EE70CB37F435EE00 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 462;
  Best Local Similarity   62.5%;  Pred. No. 70;
  Matches   10;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      11 AARAGGGCGGGGIEGP 26
           ||| ||  ||||   |
Db       2 AARRGGAAGGGGPRSP 17


RESULT  34
GATA_MYCTU
ID   GATA_MYCTU     STANDARD;      PRT;   494 AA.
AC   O53258;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT A (EC 6.3.5.-) (GLU-ADT
DE   SUBUNIT A).
GN   GATA OR RV3011C OR MT3091 OR MTV012.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC       GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
CC       TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC       REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC       AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC       + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC   -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021287; CAA16096.1; -.
DR   EMBL; AE007128; AAK47420.1; -.
DR   TIGR; MT3091; -.
DR   TubercuList; Rv3011c; -.
DR   InterPro; IPR000120; Amidase.
DR   Pfam; PF01425; Amidase; 1.
DR   PROSITE; PS00571; AMIDASES; 1.
KW   Protein biosynthesis; Ligase; Complete proteome.
FT   CONFLICT    420    420       M -> L (IN REF. 2).
SQ   SEQUENCE   494 AA;  51438 MW;  99A8824ABC436CA6 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 494;
  Best Local Similarity   52.6%;  Pred. No. 74;
  Matches   10;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      -3 GPTLRQWLAARAGGGCGGG 21
           |||   |   |  || |||
Db     141 GPTRNPWNLDRVPGGSGGG 159


RESULT  35
CG12_YEAST
ID   CG12_YEAST     STANDARD;      PRT;   545 AA.
AC   P20438;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   G1/S-SPECIFIC CYCLIN CLN2.
GN   CLN2 OR YPL256C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89345642; PubMed=2569741;
RA   Hadwiger J.A., Wittenberg C., Richardson H.E., de Barros Lopes M.,
RA   Reed S.I.;
RT   "A family of cyclin homologs that control the G1 phase in yeast.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:6255-6259(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90326560; PubMed=2197605;
RA   Hadwiger J.A., Reed S.I.;
RT   "Nucleotide sequence of the Saccharomyces cerevisiae CLN1 and CLN2
RT   genes.";
RL   Nucleic Acids Res. 18:4025-4025(1990).
RN   [3]
RP   REVISIONS.
RA   Wittenberg C., Chapman-Shimshoni D.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Messenguy F., Dubois E., Vierendeels F., Scherens B.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC       (START) TRANSITION. INTERACTS WITH THE CDC28 PROTEIN KINASE TO
CC       FORM MPF.
CC   -!- DEVELOPMENTAL STAGE: CLN1 AND CLN2 MRNAS FLUCTUATE PERIODICALLY IN
CC       THE CELL CYCLE, PEAKING IN G1 PHASE.
CC   -!- MISCELLANEOUS: A DOMINANT MUTATION IN CLN2 GENE (CLN2-1), ADVANCES
CC       THE G1- TO S-PHASE TRANSITION IN CYCLING CELLS AND IMPAIRS THE
CC       ABILITY OF CELLS TO ARREST IN G1 PHASE IN RESPONSE TO EXTERNAL
CC       SIGNALS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. STRONGEST TO OTHER
CC       G1/S CYCLINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; M33265; AAA65725.1; -.
DR   EMBL; Z73612; CAA97982.1; -.
DR   PIR; B33289; COBYC2.
DR   SGD; S0006177; CLN2.
DR   InterPro; IPR000553; Cyclin.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division.
SQ   SEQUENCE   545 AA;  61696 MW;  D6426B94E040E960 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 545;
  Best Local Similarity   38.7%;  Pred. No. 81;
  Matches   12;  Conservative    1;  Mismatches    8;  Indels   10;  Gaps    1;

Qy       9 WLAARAGGGCG----------GGGIEGPTLR 29
           ||||:  |||          ||   ||  |
Db     125 WLAAKTWGGCNHIINNVVIPTGGRFYGPNPR 155


RESULT  36
CG11_YEAST
ID   CG11_YEAST     STANDARD;      PRT;   546 AA.
AC   P20437;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   G1/S-SPECIFIC CYCLIN CLN1.
GN   CLN1 OR YMR199W OR YM9646.13.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89345642; PubMed=2569741;
RA   Hadwiger J.A., Wittenberg C., Richardson H.E., de Barros Lopes M.,
RA   Reed S.I.;
RT   "A family of cyclin homologs that control the G1 phase in yeast.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:6255-6259(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90326560; PubMed=2197605;
RA   Hadwiger J.A., Reed S.I.;
RT   "Nucleotide sequence of the Saccharomyces cerevisiae CLN1 and CLN2
RT   genes.";
RL   Nucleic Acids Res. 18:4025-4025(1990).
RN   [3]
RP   REVISIONS.
RA   Wittenberg C., Chapman-Shimshoni D.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC       (START) TRANSITION. INTERACTS WITH THE CDC28 PROTEIN KINASE TO
CC       FORM MPF.
CC   -!- DEVELOPMENTAL STAGE: CLN1 AND CLN2 MRNAS FLUCTUATE PERIODICALLY IN
CC       THE CELL CYCLE, PEAKING IN G1 PHASE.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. STRONGEST TO OTHER
CC       G1/S CYCLINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33264; AAA65724.1; -.
DR   EMBL; Z47815; CAA87822.1; -.
DR   PIR; A33289; COBYC1.
DR   SGD; S0004812; CLN1.
DR   InterPro; IPR000553; Cyclin.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division.
SQ   SEQUENCE   546 AA;  62049 MW;  4D7189B83D7A2B34 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 546;
  Best Local Similarity   38.7%;  Pred. No. 81;
  Matches   12;  Conservative    1;  Mismatches    8;  Indels   10;  Gaps    1;

Qy       9 WLAARAGGGCG----------GGGIEGPTLR 29
           ||||:  |||          ||   ||  |
Db     123 WLAAKTWGGCNHIINNVSIPTGGRFYGPNPR 153


RESULT  37
CNA1_DROME
ID   CNA1_DROME     STANDARD;      PRT;   584 AA.
AC   P12252;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CAMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17) (LEARNING/
DE   MEMORY PROCESS PROTEIN).
GN   DUNCE OR DNC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND REVISIONS.
RC   STRAIN=CANTON-S;
RX   MEDLINE=92085274; PubMed=1660926;
RA   Qiu Y.H., Chen C.-N., Malone T., Richter L., Beckendorf S.K.,
RA   Davis R.L.;
RT   "Characterization of the memory gene dunce of Drosophila
RT   melanogaster.";
RL   J. Mol. Biol. 222:553-565(1991).
RN   [2]
RP   SEQUENCE OF 223-584 FROM N.A.
RX   MEDLINE=87092243; PubMed=3025834;
RA   Chen C.-N., Denome S., Davis R.L.;
RT   "Molecular analysis of cDNA clones and the corresponding genomic
RT   coding sequences of the Drosophila dunce+ gene, the structural gene
RT   for cAMP phosphodiesterase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9313-9317(1986).
CC   -!- CATALYTIC ACTIVITY: ADENOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC       ADENOSINE 5'-PHOSPHATE.
CC   -!- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT FORMS ARE GENERATED BY THE USE OF
CC       DIFFERENT TRANSCRIPTION START SITES AND SPLICE PATTERNS.
CC   -!- DISEASE: MUTATION OF DUNCE PRODUCES FEMALE FLIES THAT ARE STERILE.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55167; CAA38960.1; -.
DR   EMBL; X55168; CAA38960.1; JOINED.
```

```
DR   EMBL; X55169; CAA38960.1; JOINED.
DR   EMBL; X55170; CAA38960.1; JOINED.
DR   EMBL; X55171; CAA38960.1; JOINED.
DR   EMBL; X55172; CAA38960.1; JOINED.
DR   EMBL; X55173; CAA38960.1; JOINED.
DR   EMBL; X55174; CAA38960.1; JOINED.
DR   EMBL; X55175; CAA38960.1; JOINED.
DR   EMBL; M14982; AAC34201.1; -.
DR   EMBL; M14978; AAC34201.1; JOINED.
DR   EMBL; M14979; AAC34201.1; JOINED.
DR   EMBL; M14980; AAC34201.1; JOINED.
DR   EMBL; M14981; AAC34201.1; JOINED.
DR   PIR; A26651; A26651.
DR   FlyBase; FBgn0000479; dnc.
DR   InterPro; IPR003607; HDc.
DR   InterPro; IPR002073; PDEase.
DR   Pfam; PF00233; PDEase; 1.
DR   PRINTS; PR00387; PDIESTERASE1.
DR   SMART; SM00471; HDc; 1.
DR   PROSITE; PS00126; PDEASE_I; 1.
KW   Hydrolase; cAMP; Alternative splicing.
FT   DOMAIN      305    310       PART OF CAMP BINDING SITE (BY SIMILARITY
FT                                TO MAMMALIAN REGULATORY SUBUNIT OF TYPE 2
FT                                CAMP DEPENDENT PROTEIN KINASE).
FT   DOMAIN      542    551       THR-RICH.
FT   DOMAIN      559    567       GLY-RICH.
SQ   SEQUENCE   584 AA;  64875 MW;  99239BE33C620501 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 584;
  Best Local Similarity   68.8%;  Pred. No. 86;
  Matches   11;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      11 AARAGGGCGGGGIEGP 26
           | ||||| ||||   |
Db     555 ALRAGGGGGGGGGMAP 570


RESULT  38
K1CJ_HUMAN
ID   K1CJ_HUMAN     STANDARD;      PRT;   593 AA.
AC   P13645;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   KERATIN, TYPE I CYTOSKELETAL 10 (CYTOKERATIN 10) (K10) (CK 10).
GN   KRT10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89125611; PubMed=2464696;
RA   Rieger M., Franke W.W.;
RT   "Identification of an orthologous mammalian cytokeratin gene. High
RT   degree of intron sequence conservation during evolution of human
RT   cytokeratin 10.";
RL   J. Mol. Biol. 204:841-856(1988).
RN   [2]
RP   SEQUENCE OF 130-593 FROM N.A.
RX   MEDLINE=88122104; PubMed=2448602;
RA   Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
RT   "Sequence of a cDNA encoding human keratin No 10 selected according
RT   to structural homologies of keratins and their tissue-specific
RT   expression.";
RL   Mol. Biol. Rep. 12:277-283(1987).
RN   [3]
RP   SEQUENCE OF 197-593 FROM N.A.
RX   MEDLINE=92339897; PubMed=1378806;
RA   Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,
RA   Kisselev L.L.;
RT   "Exons I and VII of the gene (Ker10) encoding human keratin 10
RT   undergo structural rearrangements within repeats.";
RL   Gene 116:245-251(1992).
RN   [4]
RP   SEQUENCE OF 180-184 AND 577-589.
RC   TISSUE=Keratinocytes;
RX   MEDLINE=93162043; PubMed=1286667;
RA   Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA   Vandekerckhove J.;
RT   "Microsequences of 145 proteins recorded in the two-dimensional gel
RT   protein database of normal human epidermal keratinocytes.";
RL   Electrophoresis 13:960-969(1992).
RN   [5]
RP   VARIANT EHK HIS-156.
RX   MEDLINE=92386600; PubMed=1381287;
RA   Cheng J., Syder A.J., Yu Q.-C., Letai A., Paller A.S., Fuchs E.;
RT   "The genetic basis of epidermolytic hyperkeratosis: a disorder of
RT   differentiation-specific epidermal keratin genes.";
RL   Cell 70:811-819(1992).
RN   [6]
RP   VARIANTS.
RX   MEDLINE=92141228; PubMed=1371013;
RA   Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;
RT   "Extensive size polymorphism of the human keratin 10 chain resides in
RT   the C-terminal V2 subdomain due to variable numbers and sizes of
RT   glycine loops.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
RN   [7]
RP   VARIANTS EHK HIS-156 AND SER-161.
RX   MEDLINE=92376531; PubMed=1380725;
RA   Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA   Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT   "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT   hyperkeratosis.";
RL   Science 257:1128-1130(1992).
RN   [8]
RP   VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.
RX   MEDLINE=94136477; PubMed=7508181;
RA   Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
RA   Compton J.G., Bale S.J.;
RT   "Preferential sites in keratin 10 that are mutated in epidermolytic
RT   hyperkeratosis.";
RL   Am. J. Hum. Genet. 54:179-190(1994).
RN   [9]
RP   VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.
RX   MEDLINE=94216497; PubMed=7512983;
RA   Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;
RT   "Genetic mutations in the K1 and K10 genes of patients with
RT   epidermolytic hyperkeratosis. Correlation between location and
RT   disease severity.";
RL   J. Clin. Invest. 93:1533-1542(1994).
RN   [10]
RP   VARIANT EHK ASN-160.
RX   MEDLINE=94117868; PubMed=7507150;
RA   Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;
RT   "Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene
RT   sequencing.";
RL   J. Invest. Dermatol. 102:13-16(1994).
RN   [11]
RP   VARIANTS EHK PRO-156 AND SER-156.
RX   MEDLINE=94117870; PubMed=7507152;
RA   McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,
RA   Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA   Morley S.M.;
RT   "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT   congenital ichthyosiform erythroderma (BCIE).";
RL   J. Invest. Dermatol. 102:24-30(1994).
RN   [12]
RP   VARIANT EHK THR-150.
RX   MEDLINE=95059228; PubMed=7526210;
RA   Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,
RA   Fuchs E.;
RT   "Genetic and clinical mosaicism in a type of epidermal nevus.";
RL   New Engl. J. Med. 331:1408-1415(1994).
```

```
RN   [13]
RP   VARIANT AEI THR-446.
RX   MEDLINE=99072665; PubMed=9856845;
RA   Suga Y., Duncan K.O., Heald P.W., Roop D.R.;
RT   "A novel helix termination mutation in keratin 10 in annular
RT   epidermolytic ichthyosis, a variant of bullous congenital
RT   ichthyosiform erythroderma.";
RL   J. Invest. Dermatol. 111:1220-1223(1998).
RN   [14]
RP   VARIANT EHK SER-160.
RX   MEDLINE=99215719; PubMed=10201536;
RA   Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,
RA   Hohl D., Rothnagel J.A., Roop D.R.;
RT   "A novel substitution in keratin 10 in epidermolytic hyperkeratosis.";
RL   J. Invest. Dermatol. 112:506-508(1999).
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC   -!- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
CC       STRATUM CORNEUM.
CC   -!- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
CC       THE GLY-RICH REGION (POSITIONS 490-560).
CC   -!- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC
CC       HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC       ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER
CC       CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC       CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH
CC       REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.
CC       WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION
CC       DIMINISH AND HYPERKERATOSES DEVELOP. TRANSMISSION IS AUTOSOMAL
CC       DOMINANT, BUT MOST CASES ARE SPORADIC.
CC   -!- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC
CC       ICHTHYOSIS (AEI), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC
CC       HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF
CC       BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF
CC       SIEMENS.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC       (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
CC       POSITIONS 513 TO 555.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14487; CAA32649.1; -.
DR   EMBL; M19156; AAA59468.1; -.
DR   EMBL; M77663; AAA59199.1; -.
DR   EMBL; L20218; AAB59438.1; -.
DR   EMBL; L20219; AAB59439.1; -.
DR   PIR; S02158; KRHU0.
DR   Aarhus/Ghent-2DPAGE; 7405; IEF.
DR   MIM; 148080; -.
DR   MIM; 113800; -.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR002957; Keratin_I.
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Keratin; Disease mutation;
KW   Polymorphism.
FT   DOMAIN       1    145       HEAD.
FT   DOMAIN     146    456       ROD.
FT   DOMAIN     457    593       TAIL.
FT   DOMAIN     146    181       COIL 1A.
FT   DOMAIN     182    202       LINKER 1.
FT   DOMAIN     203    294       COIL 1B.
FT   DOMAIN     295    317       LINKER 12.
FT   DOMAIN     318    456       COIL 2.
FT   DOMAIN       6    144       GLY/PHE/SER-RICH.
FT   DOMAIN     451    590       GLY/SER-RICH.
FT   VARIANT    126    126       G -> S.
FT                               /FTId=VAR_010505.
FT   VARIANT    150    150       M -> R (IN EHK).
FT                               /FTId=VAR_010506.
FT   VARIANT    150    150       M -> T (IN EHK).
FT                               /FTId=VAR_010507.
FT   VARIANT    154    154       N -> H (IN EHK).
FT                               /FTId=VAR_003826.
FT   VARIANT    156    156       R -> H (IN EHK).
FT                               /FTId=VAR_003827.
FT   VARIANT    156    156       R -> C (IN EHK).
FT                               /FTId=VAR_003828.
FT   VARIANT    156    156       R -> P (IN EHK).
FT                               /FTId=VAR_003829.
FT   VARIANT    156    156       R -> S (IN EHK).
FT                               /FTId=VAR_003830.
FT   VARIANT    160    160       Y -> D (IN EHK; SEVERE PHENOTYPE).
FT                               /FTId=VAR_003831.

  Query Match             26.4%;  Score 52;  DB 1;  Length 593;
  Best Local Similarity   47.4%;  Pred. No. 87;
  Matches    9;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 RQWLAARAGGGCGGGGIEG 25
           : : ::|:||| ||||  |
Db       9 KHYSSSRSGGGGGGGGCGG 27


RESULT  39
MYSC_ACACA
ID   MYSC_ACACA     STANDARD;      PRT;  1168 AA.
AC   P10569;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MYOSIN IC HEAVY CHAIN.
GN   MIC.
OS   Acanthamoeba castellanii (Amoeba).
OC   Eukaryota; Acanthamoebidae; Acanthamoeba.
OX   NCBI_TaxID=5755;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88016163; PubMed=3477803;
RA   Jung G., Korn E.D., Hammer J.A. III;
RT   "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RT   and non-myosin-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=86259656; PubMed=3014500;
RA   Hammer J.A. III, Jung G., Korn E.D.;
RT   "Genetic evidence that Acanthamoeba myosin I is a true myosin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
RN   [3]
RP   PHOSPHORYLATION SITE.
RX   MEDLINE=90037074; PubMed=2530230;
RA   Brzeska H., Lynch T.J., Martin B., Korn E.D.;
RT   "The localization and sequence of the phosphorylation sites of
RT   Acanthamoeba myosins I. An improved method for locating the
RT   phosphorylated amino acid.";
RL   J. Biol. Chem. 264:19340-19348(1989).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC   -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC       AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC   -!- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
CC       MYOSINS I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
CC       VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
CC       THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
CC       OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
```

```
CC   -!- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
CC       MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02974; AAA27707.1; -.
DR   PIR; A33891; MWAXIC.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT   DOMAIN        1    670       MYOSIN HEAD-LIKE.
FT   DOMAIN      671    922       TAIL HOMOLOGY REGION 1 (TH.1).
FT   DOMAIN      923    975       GLY/PRO/ALA-RICH (TH.2).
FT   DOMAIN      976   1035       SH3.
FT   DOMAIN     1036   1168       GLY/PRO/ALA-RICH (TH.2).
FT   NP_BIND     101    108       ATP (POTENTIAL).
FT   MOD_RES     311    311       PHOSPHORYLATION.
SQ   SEQUENCE   1168 AA;  127309 MW;  D07084B373A37A32 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 1168;
  Best Local Similarity   55.0%;  Pred. No. 1.5e+02;
  Matches   11;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       8 QWLAARAGGGCGGGGIEGPT 27
           | | |: ||| || |  ||:
Db     920 QILGAKGGGGGGGRGRGGPS 939


RESULT  40
NIR_NEUCR
ID   NIR_NEUCR      STANDARD;      PRT;  1176 AA.
AC   P38681;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NITRITE REDUCTASE [NAD(P)H] (EC 1.6.6.4).
GN   NIT-6.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74-OR23-1A;
RX   MEDLINE=93224461; PubMed=8096840;
RA   Exley G.E., Colandene J.D., Garrett R.H.;
RT   "Molecular cloning, characterization, and nucleotide sequence of
RT   nit-6, the structural gene for nitrite reductase in Neurospora
RT   crassa.";
RL   J. Bacteriol. 175:2379-2392(1993).
CC   -!- CATALYTIC ACTIVITY: 3 NAD(P)H + NITRITE = 3 NAD(P)(+) + NH(4)OH +
CC       H(2)O.
CC   -!- COFACTOR: THIS ENZYME IS A FAD FLAVOPROTEIN THAT ALSO CONTAINS
CC       A SIROHEME AND ONE 2FE-2S IRON-SULFUR CENTER.
CC   -!- PATHWAY: SECOND STEP IN NITRATE ASSIMILATION (DENITRIFICATION).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- INDUCTION: BY NITRATE.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC       FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC       SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07391; -; NOT_ANNOTATED_CDS.
DR   PIR; A49848; A49848.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   InterPro; IPR000660; Nir_Sir.
DR   Pfam; PF01077; NIR_SIR; 1.
DR   Pfam; PF00070; pyr_redox; 1.
DR   PRINTS; PR00397; SIROHAEM.
DR   PROSITE; PS00365; NIR_SIR; 1.
KW   Oxidoreductase; FAD; Flavoprotein; Iron-sulfur; Nitrate assimilation;
KW   Heme; NADP.
FT   NP_BIND      26     60       FAD (POTENTIAL).
FT   NP_BIND     183    215       NAD(P)H (POTENTIAL).
FT   METAL       717    717       IRON-SULFUR (2FE-2S) (POTENTIAL).
FT   METAL       723    723       IRON-SULFUR (2FE-2S) (POTENTIAL).
FT   METAL       757    757       IRON-SULFUR (2FE-2S) (POTENTIAL).
FT   METAL       761    761       IRON-SULFUR (2FE-2S) AND SIROHEME
FT                                 (BY SIMILARITY).
FT   DOMAIN      998   1054       PRO/SER-RICH.
SQ   SEQUENCE   1176 AA;  127367 MW;  FFC7DCE66F80C710 CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 1176;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 RQWLAARAGGGCGGGGIE 24
           :: : |  |||||| |:|
Db    1158 KKLVEAVGGGGCGGPGLE 1175


RESULT  41
PHYB_SORBI
ID   PHYB_SORBI     STANDARD;      PRT;  1178 AA.
AC   P93527;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PHYTOCHROME B.
GN   PHYB OR MA3.
OS   Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Sorghum.
OX   NCBI_TaxID=4558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. 58M;
RX   MEDLINE=20188796; PubMed=10723737;
RA   Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;
RT   "The phytochrome gene family in tomato and the rapid differential
RT   evolution of this family in angiosperms.";
RL   Mol. Biol. Evol. 17:362-373(2000).
RN   [2]
RP   SEQUENCE OF 208-1178 FROM N.A.
RC   STRAIN=CV. 58M;
RX   MEDLINE=97198556; PubMed=9046599;
RA   Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H.,
RA   Morgan P.W., Mullet J.E.;
```

```
RT   "The sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a
RT   phytochrome B.";
RL   Plant Physiol. 113:611-619(1997).
CC   -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC       ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC       MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC       ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
CC       PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC       RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC       RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC       GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC       BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC       PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS THE
CC       EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC   -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF182394; AAB41398.2; -.
DR   InterPro; IPR000410; Bctrl_sensor.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR003594; HATPase_c.
DR   InterPro; IPR003661; His_kinA.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR001294; Phytochrome.
DR   Pfam; PF01590; GAF; 2.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00989; PAS; 4.
DR   Pfam; PF00360; phytochrome; 2.
DR   Pfam; PF00512; signal; 2.
DR   PRINTS; PR01033; PHYTOCHROME.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00091; PAS; 2.
DR   PROSITE; PS00245; PHYTOCHROME_1; 1.
DR   PROSITE; PS50046; PHYTOCHROME_2; 1.
KW   Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW   Multigene family.
FT   DOMAIN       23     31       POLY-HIS.
FT   DOMAIN       43     54       POLY-GLY.
FT   BINDING     372    372       CHROMOPHORE (BY SIMILARITY).
SQ   SEQUENCE   1178 AA;  129136 MW;  C406DF221197B93F CRC64;


  Query Match             26.4%;  Score 52;  DB 1;  Length 1178;
  Best Local Similarity   68.8%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      12 ARAGGGCGGGGIEGPT 27
           :||||| |||| | |
Db      40 SRAGGGGGGGGGGGGT 55


RESULT  42
RU17_MOUSE
ID   RU17_MOUSE     STANDARD;      PRT;   378 AA.
AC   Q62376;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA (U1 SNRNP 70 KDA) (SNRNP70)
DE   (FRAGMENT).
GN   SNRP70.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=BALB/C;
RX   MEDLINE=89276388; PubMed=2525092;
RA   Hornig H., Fischer U., Costas M., Rauh A., Luehrmann R.;
RT   "Analysis of genomic clones of the murine U1RNA-associated 70-kDa
RT   protein reveals a high evolutionary conservation of the protein
RT   between human and mouse.";
RL   Eur. J. Biochem. 182:45-50(1989).
CC   -!- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE LOOP
CC       I REGION OF U1-SNRNA. THE TRUNCATED ISOFORM CANNOT BIND U1-SNRNA
CC       (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE C-
CC       TERMINAL REGION (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15769; CAA33777.1; -.
DR   EMBL; X15770; CAA33777.1; JOINED.
DR   EMBL; X15771; CAA33777.1; JOINED.
DR   EMBL; X15772; CAA33777.1; JOINED.
DR   EMBL; X15774; CAA33777.1; JOINED.
DR   EMBL; X15775; CAA33777.1; JOINED.
DR   EMBL; X15776; CAA33777.1; JOINED.
DR   HSSP; P09651; 1HA1.
DR   MGD; MGI:98341; Snrp70.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; Ribonucleoprotein; RNA-binding; Phosphorylation;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN       33    111       RNA-BINDING (RRM).
FT   DOMAIN      161    240       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      241    256       POLY-GLY.
FT   DOMAIN      286    333       ARG/ASP/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      334    339       POLY-GLY.
FT   VARSPLIC     90     96       AYKHADG -> TTQLACS (IN ISOFORM 2).
FT   VARSPLIC     97    378       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   378 AA;  43722 MW;  E669C31BCA365AA0 CRC64;


  Query Match             26.1%;  Score 51.5;  DB 1;  Length 378;
  Best Local Similarity   44.0%;  Pred. No. 67;
  Matches   11;  Conservative    3;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGG---GGIE 24
           | |:: |   | ||| ||   || :
Db     111 GRTVKGWRPRRLGGGLGGTRRGGAD 135


RESULT  43
RU17_HUMAN
ID   RU17_HUMAN     STANDARD;      PRT;   437 AA.
AC   P08621; P78493; Q99377; Q15364; P78494; Q15686; Q15687; Q9UE45;
AC   Q9UE46; Q9UE47; Q9UE48; Q15689; Q9UFQ6;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
```

```
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA (U1 SNRNP 70 KDA) (SNRNP70)
DE   (U1-70K).
GN   SNRP70 OR RPU1 OR U1AP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=87133480; PubMed=3028775;
RA   Theissen H., Etzerodt M., Reuter R., Schneider C., Lottspeich F.,
RA   Argos P., Luhrmann R., Philipson L.;
RT   "Cloning of the human cDNA for the U1 RNA-associated 70K protein.";
RL   EMBO J. 5:3209-3217(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=88096573; PubMed=2447561;
RA   Spritz R.A., Strunk K., Surowy C.S., Hoch S.O., Barton D.E.,
RA   Francke U.;
RT   "The human U1-70K snRNP protein: cDNA cloning, chromosomal
RT   localization, expression, alternative splicing and RNA-binding.";
RL   Nucleic Acids Res. 15:10373-10391(1987).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND RNA-BINDING DOMAIN.
RX   MEDLINE=89195226; PubMed=2467746;
RA   Query C.C., Bentley R.C., Keene J.D.;
RT   "A common RNA recognition motif identified within a defined U1 RNA
RT   binding domain of the 70K U1 snRNP protein.";
RL   Cell 57:89-101(1989).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=91065657; PubMed=2147422;
RA   Spritz R.A., Strunk K., Surowy C.S., Mohrenweiser H.W.;
RT   "Human U1-70K ribonucleoprotein antigen gene: organization, nucleotide
RT   sequence, and mapping to locus 19q13.3.";
RL   Genomics 8:371-379(1990).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Liver;
RX   MEDLINE=96363460; PubMed=8746626;
RA   Northemann W., Berg H., Stahnke G., Walter M., Hunt N., Fenning S.;
RT   "Identification of an inhibitory element within the human 68-kDa (U1)
RT   ribonucleoprotein antigen.";
RL   Protein Expr. Purif. 6:748-756(1995).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM 4).
RC   TISSUE=Testis;
RA   Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 219-348, AND PHOSPHORYLATION.
RX   MEDLINE=93324364; PubMed=8332490;
RA   Woppmann A., Will C.L., Kornstaedt U., Zuo P., Manley J.L.,
RA   Luehrmann R.;
RT   "Identification of an snRNP-associated kinase activity that
RT   phosphorylates arginine/serine rich domains typical of splicing
RT   factors.";
RL   Nucleic Acids Res. 21:2815-2822(1993).
CC   -!- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE LOOP
CC       I REGION OF U1-SNRNA. THE TRUNCATED ISOFORMS CANNOT BIND U1-SNRNA.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE C-
CC       TERMINAL REGION.
CC   -!- DISEASE: MAJOR RIBONUCLEOPROTEIN ANTIGEN RECOGNIZED BY THE SERA
CC       FROM PATIENTS WITH AUTOIMMUNE DISEASES, SUCH AS SYSTEMIC LUPUS
CC       ERYTHEMATOSUS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04654; CAA28352.1; ALT_INIT.
DR   EMBL; X06814; CAA29963.1; -.
DR   EMBL; X06815; CAA29964.1; ALT_INIT.
DR   EMBL; X06817; CAA29966.1; ALT_INIT.
DR   EMBL; X07402; CAA30304.1; -.
DR   EMBL; X06816; CAA29965.1; -.
DR   EMBL; X06812; CAA29961.1; -.
DR   EMBL; X06811; CAA29960.1; -.
DR   EMBL; X07401; CAA30303.1; -.
DR   EMBL; X07403; CAA30305.1; -.
DR   EMBL; M22636; AAA03001.1; -.
DR   EMBL; M57939; AAA36572.1; -.
DR   EMBL; M57929; AAA36572.1; JOINED.
DR   EMBL; M57930; AAA36572.1; JOINED.
DR   EMBL; M57932; AAA36572.1; JOINED.
DR   EMBL; M57934; AAA36572.1; JOINED.
DR   EMBL; M57937; AAA36572.1; JOINED.
DR   EMBL; M57939; AAA36573.1; -.
DR   EMBL; M57929; AAA36573.1; JOINED.
DR   EMBL; M57930; AAA36573.1; JOINED.
DR   EMBL; M57932; AAA36573.1; JOINED.
DR   EMBL; M57934; AAA36573.1; JOINED.
DR   EMBL; M57937; AAA36573.1; JOINED.
DR   EMBL; M57935; AAA36571.1; -.
DR   EMBL; M57929; AAA36571.1; JOINED.
DR   EMBL; M57930; AAA36571.1; JOINED.
DR   EMBL; M57932; AAA36571.1; JOINED.
DR   EMBL; M57934; AAA36571.1; JOINED.
DR   EMBL; X84841; CAA59278.1; -.
DR   EMBL; AL117507; CAB55969.1; -.
DR   PIR; A25707; A25707.
DR   HSSP; P09651; 1UP1.
DR   MIM; 180740; -.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing;
KW   Phosphorylation; Alternative splicing.
FT   DOMAIN      103    181       RNA-BINDING (RRM).
FT   DOMAIN      231    310       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      311    314       POLY-GLY.
FT   DOMAIN      344    393       ARG/ASP/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      394    398       POLY-GLY.
FT   VARSPLIC      1    159       MTQFLPPNLLALFAPRDPIPYLPPLEKLPHEKHHNQPYCGI
FT                                APYIREFEDPRDAPPPTRAETREERMERKRREKIERRQQEV
FT                                ETELKMWDPHNDPNAQGDAFKTLFVARVNYDTTESKLRREF
FT                                EVYGPIKRIHMVYSKRSGKPRGYAFIEYEHERDMHS -> M
FT                                EQALHRFGRGLVWLSVAWLSVGRVRVRDDGDTGRGFCRAGP
FT                                VLTRGPSGDSSPLPLPTSVTA (IN ISOFORM 4).
FT   VARSPLIC    160    166       AYKHADG -> TTQLACS (IN ISOFORM 3).
FT   VARSPLIC    167    437       MISSING (IN ISOFORM 3).
FT   VARSPLIC    223    231       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   437 AA;  51556 MW;  F1020BF5C40CF97D CRC64;


  Query Match             26.1%;  Score 51.5;  DB 1;  Length 437;
  Best Local Similarity   44.0%;  Pred. No. 76;
  Matches   11;  Conservative    3;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGG---GGIE 24
           | |:: |   | ||| ||   || :
Db     181 GRTVKGWRPRRLGGGLGGTRRGGAD 205


RESULT  44
FXGA_CHICK
```

```
ID   FXGA_CHICK     STANDARD;      PRT;   440 AA.
AC   Q98937;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FORKHEAD PROTEIN G1A (FORKHEAD-RELATED PROTEIN FKHL2) (TRANSCRIPTION
DE   FACTOR BF-2) (BRAIN FACTOR 2) (BF2) (CBF-2) (T-14-6).
GN   FOXG1A OR FKHL2 OR HFHBF2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Retina;
RX   MEDLINE=96338226; PubMed=8757134;
RA   Yuasa J., Hirano S., Yamagata M., Noda M.;
RT   "Visual projection map specified by topographic expression of
RT   transcription factors in the retina.";
RL   Nature 382:632-635(1996).
CC   -!- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND
CC       CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL
CC       GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF
CC       THEIR TARGET GENES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: RETINA.
CC   -!- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING
CC       PRIMORDIAL RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT
CC       E3, EXPRESSED IN THE TEMPORAL RETINA AND ASSOCIATED PIGMENT
CC       EPITHELIUM AS WELL AS IN PART OF THE DIENCEPHALON, AND AT E7 IS
CC       EXPRESSED IN RETINAL GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM
CC       E4 AND ALMOST DISAPPEAR BY E10.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47276; AAB08467.1; -.
DR   InterPro; IPR001766; Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Developmental protein.
FT   DOMAIN       39     43       POLY-ARG.
FT   DOMAIN       49     58       POLY-GLU.
FT   DOMAIN       77     88       POLY-GLY.
FT   DOMAIN      109    112       POLY-ALA.
FT   DOMAIN      113    137       POLY-GLY.
FT   DNA_BIND    143    237       FORK-HEAD.
FT   DOMAIN      321    330       POLY-PRO.
SQ   SEQUENCE   440 AA;  44669 MW;  82266C2E1E103A48 CRC64;


  Query Match             26.1%;  Score 51.5;  DB 1;  Length 440;
  Best Local Similarity   60.0%;  Pred. No. 76;
  Matches   12;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGG 22
           ||   |  ||  ||| ||||
Db     105 GP---QKAAAAGGGGAGGGG 121


RESULT  45
RU17_XENLA
ID   RU17_XENLA     STANDARD;      PRT;   471 AA.
AC   P09406;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA (U1 SNRNP 70 KDA) (SNRNP70)
DE   (U1-70K).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89210819; PubMed=2468488;
RA   Etzerodt M., Vignali R., Scherly D., Mattaj I.W., Ciliberto G.,
RA   Philipson L.;
RT   "Structure and expression of a Xenopus gene encoding an snRNP protein
RT   (U1 70K).";
RL   EMBO J. 7:4311-4321(1988).
CC   -!- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE STEM
CC       LOOP I REGION OF U1-SNRNA.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12430; CAA30972.1; -.
DR   PIR; S02016; S02016.
DR   HSSP; P09651; 1UP1.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing.
FT   DOMAIN      103    184       RNA-BINDING (RRM).
FT   DOMAIN      221    320       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      344    426       ARG/ASP/GLU-RICH (MIXED CHARGE).
SQ   SEQUENCE   471 AA;  57203 MW;  959A8BB3064C3FC3 CRC64;


  Query Match             26.1%;  Score 51.5;  DB 1;  Length 471;
  Best Local Similarity   44.0%;  Pred. No. 81;
  Matches   11;  Conservative    3;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGG---GGIE 24
           | |:: |   | ||| ||   || :
Db     184 GRTVKGWRPRRLGGGLGGTRRGGAD 208


Search completed: December 26, 2001, 10:29:29
Job time: 95 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:27:23 ; Search time 22.51 Seconds
(without alignments)
233.932 Million cell updates/sec

Title: US-09-422-838C-33
Perfect score: 197
Sequence: 1 IEGPTLRQWLAARAGGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	67	34.0	865	2	O54108	O54108 streptomyce
2	65	33.0	360	10	Q9LGC9	Q9lgc9 oryza sativ
3	64.5	32.7	431	13	Q9PVG9	Q9pvg9 coturnix co
4	61.5	31.2	202	10	Q9FTZ5	Q9ftz5 oryza sativ
5	61	31.0	439	10	Q9SDK6	Q9sdk6 oryza sativ
6	60	30.5	464	4	Q9UEA1	Q9uea1 homo sapien
7	60	30.5	492	4	Q9UNW9	Q9unw9 homo sapien
8	60	30.5	498	4	O43267	O43267 homo sapien
9	60	30.5	500	5	Q19476	Q19476 caenorhabdi
10	60	30.5	654	5	Q9VBC7	Q9vbc7 drosophila
11	60	30.5	654	5	Q9UAE7	Q9uae7 drosophila
12	59.5	30.2	454	4	Q14060	Q14060 homo sapien
13	59.5	30.2	488	2	Q9CCC0	Q9ccc0 mycobacteri
14	59.5	30.2	496	2	Q9AD76	Q9ad76 streptomyce
15	59.5	30.2	518	2	Q49843	Q49843 mycobacteri
16	59	29.9	125	10	Q9LWC8	Q9lwc8 oryza sativ
17	59	29.9	492	11	O35392	O35392 mus musculu
18	59	29.9	683	2	O83436	O83436 treponema p
19	59	29.9	801	3	Q9HEA4	Q9hea4 neurospora
20	58.5	29.7	805	4	O95692	O95692 homo sapien
21	58.5	29.7	1431	11	Q9JMH4	Q9jmh4 mesocricetu
22	58	29.4	117	10	Q9FU26	Q9fu26 oryza sativ
23	58	29.4	134	2	Q56434	Q56434 thermus aqu
24	58	29.4	170	5	Q9W033	Q9w033 drosophila
25	58	29.4	302	3	Q99034	Q99034 trichoderma
26	58	29.4	516	10	Q9XEJ0	Q9xej0 zea mays (m
27	57.5	29.2	244	11	Q9D384	Q9d384 mus musculu
28	57.5	29.2	302	2	Q9S596	Q9s596 myxococcus
29	57.5	29.2	495	2	O33230	O33230 mycobacteri
30	57	28.9	76	10	Q9C7W8	Q9c7w8 arabidopsis
31	57	28.9	377	13	Q9YHD0	Q9yhd0 petromyzon
32	57	28.9	414	3	Q9HFM0	Q9hfm0 metarhizium
33	57	28.9	524	4	Q9BZE0	Q9bze0 homo sapien
34	57	28.9	529	10	Q9ASE5	Q9ase5 oryza sativ
35	57	28.9	607	2	Q9L8D4	Q918d4 polyangium
36	57	28.9	612	4	Q9P270	Q9p270 homo sapien
37	57	28.9	651	10	Q9LGW5	Q9lgw5 oryza sativ
38	57	28.9	1130	4	O75182	O75182 homo sapien
39	56.5	28.7	176	1	Q9YDB1	Q9ydb1 aeropyrum p
40	56.5	28.7	243	10	Q9AR44	Q9ar44 oryza sativ
41	56.5	28.7	1548	4	Q9NYW9	Q9nyw9 homo sapien
42	56.5	28.7	2161	4	Q9Y566	Q9y566 homo sapien
43	56	28.4	56	2	O34781	O34781 bacillus su
44	56	28.4	56	9	O64033	O64033 bacteriopha
45	56	28.4	163	5	O61832	O61832 caenorhabdi
46	56	28.4	349	10	Q9C7F3	Q9c7f3 arabidopsis
47	56	28.4	424	10	Q9FEB6	Q9feb6 oryza sativ
48	56	28.4	447	13	O73628	O73628 anolis caro
49	56	28.4	452	5	Q9VJK4	Q9vjk4 drosophila
50	56	28.4	540	3	Q09431	Q09431 phanerochae
51	56	28.4	767	2	O53435	O53435 mycobacteri
52	56	28.4	995	5	Q9V7E7	Q9v7e7 drosophila
53	55.5	28.2	775	4	Q9C0I1	Q9c0i1 homo sapien
54	55.5	28.2	873	10	Q9XF26	Q9xf26 oryza sativ
55	55	27.9	77	2	Q9L5N0	Q915n0 salmonella
56	55	27.9	180	3	Q9P639	Q9p639 neurospora
57	55	27.9	201	3	Q9P553	Q9p553 neurospora
58	55	27.9	257	10	O22131	O22131 arabidopsis
59	55	27.9	309	5	Q9VV01	Q9vv01 drosophila
60	55	27.9	331	5	Q9U2I1	Q9u2i1 caenorhabdi
61	55	27.9	333	5	Q9U2I0	Q9u2i0 caenorhabdi
62	55	27.9	393	5	Q18880	Q18880 caenorhabdi
63	55	27.9	399	10	Q9LDW5	Q91dw5 arabidopsis
64	55	27.9	422	5	O96755	O96755 branchiosto
65	55	27.9	517	3	Q9Y722	Q9y722 irpex lacte
66	55	27.9	556	5	Q9VLB8	Q9vlb8 drosophila
67	55	27.9	694	4	Q9H461	Q9h461 homo sapien
68	55	27.9	1024	5	Q9VFM5	Q9vfm5 drosophila
69	55	27.9	1475	10	Q9XEP3	Q9xep3 sorghum bic
70	55	27.9	2904	11	Q9EPN0	Q9epn0 mus musculu
71	55	27.9	2931	11	Q9EPM9	Q9epm9 mus musculu
72	55	27.9	2936	11	Q9EPN1	Q9epn1 mus musculu
73	54.5	27.7	246	4	Q16560	Q16560 homo sapien
74	54.5	27.7	262	12	Q9ICS7	Q9ics7 pseudorabie
75	54.5	27.7	392	10	Q9ZRB9	Q9zrb9 lycopersico
76	54.5	27.7	394	4	Q9BSE2	Q9bse2 homo sapien
77	54.5	27.7	407	2	Q9L0B6	Q910b6 streptomyce
78	54.5	27.7	453	5	Q9NGF7	Q9ngf7 drosophila
79	54.5	27.7	453	5	Q9NGF6	Q9ngf6 drosophila
80	54.5	27.7	453	5	Q9N6M8	Q9n6m8 drosophila
81	54.5	27.7	584	10	Q9LI16	Q91i16 oryza sativ
82	54.5	27.7	850	5	Q9W4F0	Q9w4f0 drosophila
83	54.5	27.7	1028	5	Q9W4F1	Q9w4f1 drosophila
84	54	27.4	137	10	Q9M6A1	Q9m6a1 catharanthu
85	54	27.4	139	5	Q9W2W0	Q9w2w0 drosophila
86	54	27.4	160	10	Q9M699	Q9m699 catharanthu
87	54	27.4	175	10	Q9LRR3	Q91rr3 arabidopsis
88	54	27.4	296	12	Q69118	Q69118 human herpe
89	54	27.4	490	10	O04270	O04270 chlamydomon
90	54	27.4	495	2	O53325	O53325 mycobacteri
91	54	27.4	665	2	Q48373	Q48373 janthinobac
92	54	27.4	688	4	Q9BYD8	Q9byd8 homo sapien

93	54	27.4	743	5	Q9VBW6	Q9vbw6 drosophila
94	54	27.4	841	10	Q9SXI9	Q9sxi9 oryza sativ
95	54	27.4	975	5	Q9V410	Q9v410 drosophila
96	54	27.4	2274	5	Q9VYU0	Q9vyu0 drosophila
97	54	27.4	2638	2	O30914	O30914 streptomyce
98	53.5	27.2	201	2	Q9Z3X4	Q9z3x4 ralstonia s
99	53.5	27.2	252	11	Q9CXS4	Q9cxs4 mus musculu
100	53.5	27.2	395	11	Q9Z0T7	Q9z0t7 rattus norv

ALIGNMENTS

```
RESULT   1
O54108
ID   O54108      PRELIMINARY;      PRT;   865 AA.
AC   O54108;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE SECRETED PROTEASE.
GN   SC10A5.17.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Murphy L., Harris D.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL021529; CAA16449.1; -.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   InterPro; IPR000601; PKD_domain.
DR   InterPro; IPR002169; Micollptase.
DR   Pfam; PF00801; PKD; 1.
DR   Pfam; PF01752; Peptidase_M9; 1.
DR   PRINTS; PR00931; MICOLLPTASE.
DR   PROSITE; PS50093; PKD; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00089; PKD; 1.
KW   Protease.
SQ   SEQUENCE   865 AA;  92392 MW;  2145740361275F8F CRC64;
```

```
  Query Match             34.0%;  Score 67;  DB 2;  Length 865;
  Best Local Similarity   66.7%;  Pred. No. 4.9;
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       9 WLAARAGGGCGGGGIEGP 26
           |||| | | |||||   |
Db     651 WLAACAAGNCGGGGTNPP 668
```

```
RESULT   2
Q9LGC9
ID   Q9LGC9      PRELIMINARY;      PRT;   360 AA.
AC   Q9LGC9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE ZINC FINGER PROTEIN.
GN   P0462H08.19.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0462H08.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002525; BAB07996.1; -.
DR   InterPro; IPR000571; Zf-CCCH.
DR   Pfam; PF00642; zf-CCCH; 4.
DR   SMART; SM00356; ZnF_C3H1; 4.
SQ   SEQUENCE   360 AA;  37368 MW;  5105598D7E1C77B2 CRC64;
```

```
  Query Match             33.0%;  Score 65;  DB 10;  Length 360;
  Best Local Similarity   52.0%;  Pred. No. 3.5;
  Matches   13;  Conservative    2;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEG 25
           :|||  |  |    ||| ||||  :|
Db      26 LEGPMWRMGLGGGGGGGGGGGGGDG 50
```

```
RESULT   3
Q9PVG9
ID   Q9PVG9      PRELIMINARY;      PRT;   431 AA.
AC   Q9PVG9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   POU-BOX PROTEIN BRAIN-2.
OS   Coturnix coturnix japonica (Japanese quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=93934;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Xue J.X., Zhang W., Fu D.C., He R.Q., Xue Z.G.;
RT   "qBrain-2, a POU-box gene expressed in quail embryos.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER HOMEOBOX DOMAINS.
DR   EMBL; AF091043; AAF00040.1; -.
DR   HSSP; P14859; 1OCT.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000327; POU.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00157; pou; 1.
DR   PRINTS; PR00028; POUDOMAIN.
DR   ProDom; PD000583; POU; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00352; POU; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PROSITE; PS00035; POU_1; 1.
DR   PROSITE; PS00465; POU_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
SQ   SEQUENCE   431 AA;  43722 MW;  1DC47E53F9ACC7D5 CRC64;
```

```
  Query Match             32.7%;  Score 64.5;  DB 13;  Length 431;
  Best Local Similarity   40.5%;  Pred. No. 4.8;
  Matches   17;  Conservative    2;  Mismatches    6;  Indels   17;  Gaps    2;
```

```
Qy       8 QWLAARA------------GGGCGGGGIEGPTLRQWLAARA 36
           ||:|| :            ||| |||| | |    | || |
Db      58 QWIAALSHGGPGGGGGGGGGGGGGGGGGGEAP----WAAAAA 95
```

RESULT 4
Q9FTZ5
ID Q9FTZ5 PRELIMINARY; PRT; 202 AA.
AC Q9FTZ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0436E04.1 PROTEIN.
GN P0436E04.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0436E04.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP002818; BAB16319.1; -.
SQ SEQUENCE 202 AA; 19763 MW; BFC2520037F8E274 CRC64;

Query Match 31.2%; Score 61.5; DB 10; Length 202;
Best Local Similarity 38.1%; Pred. No. 5;
Matches 16; Conservative 5; Mismatches 12; Indels 9; Gaps 2;

```
Qy       1 IEGPTLRQWLAARAGGGCGG--------GGIEGPTLRQWLAA 34
           :  |: |:  | |  |||||        || :|   |:| ||
Db      94 VVSPSCRRQTAGR-HGGCGGGGRWMAAAGGRDGGGCRRWWAA 134
```

RESULT 5
Q9SDK6
ID Q9SDK6 PRELIMINARY; PRT; 439 AA.
AC Q9SDK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0705D01.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP000492; BAA84610.1; -.
SQ SEQUENCE 439 AA; 47297 MW; 533EEC240CEA1BA2 CRC64;

Query Match 31.0%; Score 61; DB 10; Length 439;
Best Local Similarity 32.0%; Pred. No. 12;
Matches 16; Conservative 2; Mismatches 18; Indels 14; Gaps 1;

```
Qy       1 IEGPTLRQWLAARAGGCGGGG-------------IEGPTLRQWLAARA 36
           :  | || |     ||| ||||             : |   |   || |
Db      39 LHAPLLRLWPLGGGGGGGGGGGGGGGGERVGAVGGAVRGEEARSQRAAEA 88
```

RESULT 6
Q9UEA1
ID Q9UEA1 PRELIMINARY; PRT; 464 AA.
AC Q9UEA1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE RNA-BINDING PROTEIN NOVA-2 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhart-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT D19S412.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC006540; AAD13116.1; -.
DR InterPro; IPR000958; KH.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00013; KH-domain; 3.
DR PRINTS; PR00456; RIBOSOMALP2.
DR SMART; SM00322; KH; 3.
FT NON_TER 1 1
SQ SEQUENCE 464 AA; 45901 MW; 0B16BAE99C271CC3 CRC64;

Query Match 30.5%; Score 60; DB 4; Length 464;
Best Local Similarity 53.6%; Pred. No. 17;
Matches 15; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

```
Qy       9 WLAARAGGGCGGGGIEGPTLRQWLAARA 36
           :| | |||| ||||  || :    || |
Db     337 YLGAGAGGGAGGGG--GPLVAAAAAAGA 362
```

RESULT 7
Q9UNW9
ID Q9UNW9 PRELIMINARY; PRT; 492 AA.
AC Q9UNW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE RNA-BINDING PROTEIN NOVA-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99007301; PubMed=9789075;
RA Yang Y.Y., Yin G.L., Darnell R.B.;
RT "The neuronal RNA-binding protein Nova-2 is implicated as the
RT autoantigen targeted in POMA patients with dementia.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:13254-13259(1998).
DR EMBL; AF083898; AAC72355.1; -.
DR InterPro; IPR000958; KH.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
SQ SEQUENCE 492 AA; 49008 MW; 41B63EAF6899256B CRC64;

Query Match 30.5%; Score 60; DB 4; Length 492;
Best Local Similarity 53.6%; Pred. No. 18;
Matches 15; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

```
Qy       9 WLAARAGGGCGGGGIEGPTLRQWLAARA 36
           :| | |||| |||| || :    || |
Db     365 YLGAGAGGGAGGGG--GPLVAAAAAAGA 390


RESULT   8
O43267
ID   O43267      PRELIMINARY;      PRT;   498 AA.
AC   O43267;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ASTROCYTIC NOVA1-LIKE RNA-BINDING PROTEIN (NEUROONCOLOGIC VENTRAL
DE   ANTIGEN 3) (FRAGMENT).
GN   ANOVA OR NOVA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20197319; PubMed=10735272;
RA   Ueki K., Ramaswamy S., Billings S.J., Mohrenweiser H.W., Louis D.N.;
RT   "ANOVA, a putative astrocytic RNA binding protein gene that maps to
RT   chromosome 19q13.3.";
RL   Neurogenetics 1:31-36(1997).
CC   -!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC
CC       SUBSET OF DEVELOPING NEURONS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: BRAIN. EXPRESSION RESTRICTED TO ASTROCYTES.
CC   -!- DISEASE: MAY BE A TARGET ANTIGEN IN ONE OF THE UNDEFINED HUMAN
CC       PARANEOPLASTIC SYNDROMES.
DR   EMBL; U70477; AAB88661.1; -.
DR   InterPro; IPR000958; KH.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
KW   Nuclear protein; RNA-binding; Repeat; Antigen.
FT   NON_TER       1      1
FT   DOMAIN       42     76       KH.
FT   DOMAIN      140    174       KH.
FT   DOMAIN      246    255       POLY-ALA.
FT   DOMAIN      325    330       POLY-ALA.
FT   DOMAIN      350    356       POLY-PRO.
FT   DOMAIN      363    368       POLY-ALA.
FT   DOMAIN      375    385       POLY-GLY.
FT   DOMAIN      389    397       POLY-ALA.
FT   DOMAIN      416    446       KH.
SQ   SEQUENCE   498 AA;  49721 MW;  C4B54196FDB6BF78 CRC64;


  Query Match             30.5%;  Score 60;  DB 4;  Length 498;
  Best Local Similarity   53.6%;  Pred. No. 19;
  Matches   15;  Conservative    2;  Mismatches    9;  Indels    2;  Gaps    1;

Qy       9 WLAARAGGGCGGGGIEGPTLRQWLAARA 36
           :| | |||| |||| || :    || |
Db     371 YLGAGAGGGAGGGG--GPLVAAAAAAGA 396


RESULT   9
Q19476
ID   Q19476      PRELIMINARY;   PRT;   500 AA.
AC   Q19476;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F15B9.5 PROTEIN.
GN   F15B9.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Percy C.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z78013; CAB01420.1; -.
DR   InterPro; IPR001254; Trypsin.
SQ   SEQUENCE   500 AA;  53946 MW;  1416327086FE7CF6 CRC64;


  Query Match             30.5%;  Score 60;  DB 5;  Length 500;
  Best Local Similarity   52.2%;  Pred. No. 19;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 GPTLRQWLAARAGGGCGGGGIEG 25
           |  | ::|: | ||| ||||: |
Db     429 GSMLGRFLSNRGGGGGGGGGMGG 451


RESULT  10
Q9VBC7
ID   Q9VBC7      PRELIMINARY;      PRT;   654 AA.
AC   Q9VBC7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   AMON PROTEIN.
GN   AMON OR CG6438.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
```

```
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003757; AAF56615.1; -.
DR   HSSP; P00782; 2SBT.
DR   FlyBase; FBgn0023179; amon.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR002884; P_domain.
DR   Pfam; PF01483; P; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   ProDom; PD000717; P_domain; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
SQ   SEQUENCE   654 AA;  71733 MW;  2BD57F683929D237 CRC64;


  Query Match             30.5%;  Score 60;  DB 5;  Length 654;
  Best Local Similarity   48.0%;  Pred. No. 24;
  Matches   12;  Conservative    3;  Mismatches    8;  Indels    2;  Gaps    1;

Qy       6 LRQWLAARAGGGCG--GGGIEGPTL 28
           |  | :| |||| |  | |: || :
Db      15 LLHWASAGAGGGAGGSGAGLSGPAV 39


RESULT  11
Q9UAE7
ID   Q9UAE7      PRELIMINARY;      PRT;   654 AA.
AC   Q9UAE7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROHORMONE AND NEUROPEPTIDE PROCESSING PROTEASE.
GN   AMON OR PC2 OR CG6438.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99365357; PubMed=10436051;
RA   Siekhaus D.E., Fuller R.S.;
RT   "A role for amontillado, the Drosophila homolog of the neuropeptide
RT   precursor processing protease PC2, in triggering hatching behavior.";
RL   J. Neurosci. 19:6942-6954(1999).
DR   EMBL; AF033117; AAD49105.1; -.
DR   HSSP; P00782; 2SBT.
DR   FlyBase; FBgn0023179; amon.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR002884; P_domain.
DR   Pfam; PF01483; P; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   ProDom; PD000717; P_domain; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Protease; Neuropeptide.
SQ   SEQUENCE   654 AA;  71733 MW;  D021D4882293C996 CRC64;


  Query Match             30.5%;  Score 60;  DB 5;  Length 654;
  Best Local Similarity   48.0%;  Pred. No. 24;
  Matches   12;  Conservative    3;  Mismatches    8;  Indels    2;  Gaps    1;

Qy       6 LRQWLAARAGGGCG--GGGIEGPTL 28
           |  | :| |||| |  | |: || :
Db      15 LLHWASAGAGGGAGGSGAGLSGPAV 39


RESULT  12
Q14060
ID   Q14060      PRELIMINARY;      PRT;   454 AA.
AC   Q14060;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COPROPORPHYRINOGEN OXIDASE.
GN   CPX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=95078835; PubMed=7987309;
RA   Delfau-Larue M., Martasek P., Grandchamp B.;
RT   "Coproporphyrinogen oxidase: gene organization and description of a
RT   mutation leading to exon 6 skipping.";
RL   Hum. Mol. Genet. 3:1325-1330(1994).
RN   [2]
RP   SEQUENCE OF 101-454 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=94114558; PubMed=8286403;
RA   Taketani S., Kohno H., Furukawa T., Yoshinaga T., Tokunaga R.;
RT   "Molecular cloning, sequencing and expression of cDNA encoding human
RT   coproporphyrinogen oxidase.";
RL   Biochim. Biophys. Acta 1183:547-549(1994).
DR   EMBL; Z34531; CAA84292.1; -.
DR   EMBL; Z34803; CAA84292.1; JOINED.
DR   EMBL; Z34804; CAA84292.1; JOINED.
DR   EMBL; Z34805; CAA84292.1; JOINED.
DR   EMBL; Z34806; CAA84292.1; JOINED.
DR   EMBL; Z34807; CAA84292.1; JOINED.
DR   EMBL; Z34808; CAA84292.1; JOINED.
DR   EMBL; D16611; BAA04033.1; -.
DR   InterPro; IPR001260; Coprogen_oxidas.
DR   Pfam; PF01218; Coprogen_oxidas; 1.
DR   PRINTS; PR00073; COPRGNOXDASE.
DR   PROSITE; PS01021; COPROGEN_OXIDASE; 1.
SQ   SEQUENCE   454 AA;  50152 MW;  6EC3D15FD8FD86B5 CRC64;


  Query Match             30.2%;  Score 59.5;  DB 4;  Length 454;
  Best Local Similarity   44.4%;  Pred. No. 19;
  Matches   16;  Conservative    0;  Mismatches    5;  Indels   15;  Gaps    3;

Qy       3 GPTLRQWLAARAGGG-------CGGGGIEGPTLRQW 31
             ||    || || | |       |||||     || |
Db      11 GPC---WLVARGGCGGPRAWSQCGGGG-----LRAW 38


RESULT  13
Q9CCC0
ID   Q9CCC0      PRELIMINARY;      PRT;   488 AA.
AC   Q9CCC0;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   POSSIBLE ATP/GTP-BINDING PROTEIN.
GN   ML0997.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; AL583920; CAC31378.1; -.
DR   InterPro; IPR000765; GTP1_OBG.
DR   PRINTS; PR00326; GTP1OBG.
KW   Complete proteome.
SQ   SEQUENCE   488 AA;  52800 MW;  188918856F9774AA CRC64;


  Query Match             30.2%;  Score 59.5;  DB 2;  Length 488;
  Best Local Similarity   43.3%;  Pred. No. 21;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    7;  Gaps    1;

Qy       4 PTLRQW-------LAARAGGGCGGGGIEGP 26
           | || |       :  ||||  || |: ||
Db     189 PRLRGWGESMSRQVGGRAGGSGGGVGLRGP 218


RESULT  14
Q9AD76
ID   Q9AD76      PRELIMINARY;      PRT;   496 AA.
AC   Q9AD76;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN   SCK13.27.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL512667; CAC21636.2; -.
SQ   SEQUENCE   496 AA;  49548 MW;  54E110C4F86231A4 CRC64;


  Query Match             30.2%;  Score 59.5;  DB 2;  Length 496;
  Best Local Similarity   43.8%;  Pred. No. 21;
  Matches   14;  Conservative    3;  Mismatches    6;  Indels    9;  Gaps    1;

Qy       4 PTLRQWL---------AARAGGGCGGGGIEGP 26
             |||:  |         |  :||| ||||: ||
Db     408 PTLQAQLGGGAGGGGGAGGSGGGFGGGGLGGP 439


RESULT  15
Q49843
ID   Q49843      PRELIMINARY;      PRT;   518 AA.
AC   Q49843;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HFLX.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Robison K.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Smith D.R.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Robison K.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U00019; AAA17274.1; -.
SQ   SEQUENCE   518 AA;  56001 MW;  6641916CC84F374B CRC64;


  Query Match             30.2%;  Score 59.5;  DB 2;  Length 518;
  Best Local Similarity   43.3%;  Pred. No. 22;
  Matches   13;  Conservative    2;  Mismatches    8;  Indels    7;  Gaps    1;

Qy       4 PTLRQW-------LAARAGGGCGGGGIEGP 26
           | || |       :  ||||  || |: ||
Db     219 PRLRGWGESMSRQVGGRAGGSGGGVGLRGP 248


RESULT  16
Q9LWC8
ID   Q9LWC8      PRELIMINARY;      PRT;   125 AA.
AC   Q9LWC8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0483F08.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002094; BAA96216.1; -.
SQ   SEQUENCE   125 AA;  13396 MW;  C609D8D0B07BC505 CRC64;
```

```
  Query Match             29.9%;  Score 59;  DB 10;  Length 125;
  Best Local Similarity   40.5%;  Pred. No. 6.1;
  Matches   17;  Conservative    2;  Mismatches    9;  Indels   14;  Gaps    2;

Qy       2 EGPTLRQWLAARA----------GGGCGGGGIEGPTLRQ 30
           ||   | | |||:            |||  |||| | |  |:
Db      83 EGAAAR-WRAARSPARGGQRGGHRRRGGGGGGGGRERPRRRR 123


RESULT  17
O35392
ID   O35392      PRELIMINARY;      PRT;   492 AA.
AC   O35392;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MESODERM/MESENCHYME FORK HEAD 2.
GN   FOXD2 OR MF2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MESENCHYME;
RA   Wu S.C.-Y., Grindley J., Winnier G.E., Hargett L., Hogan B.L.H.;
RL   Mech. Dev. 0:0-0(1997).
DR   EMBL; AF023915; AAB81275.1; -.
DR   HSSP; Q63245; 2HFH.
DR   TRANSFAC; T02492; -.
DR   MGD; MGI:1347471; Foxd2.
DR   InterPro; IPR001766; Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
SQ   SEQUENCE   492 AA;  48936 MW;  7F82440F4C435702 CRC64;


  Query Match             29.9%;  Score 59;  DB 11;  Length 492;
  Best Local Similarity   70.6%;  Pred. No. 24;
  Matches   12;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 LRQWLAARAGGGCGGGG 22
           ||| |   |||| ||||
Db     384 LRQGLKTDAGGGAGGGG 400


RESULT  18
O83436
ID   O83436      PRELIMINARY;      PRT;   683 AA.
AC   O83436;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   TP0421.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AE001220; AAC65409.1; -.
DR   TIGR; TP0421; -.
DR   InterPro; IPR001258; NHL.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF01436; NHL; 4.
DR   Pfam; PF00515; TPR; 1.
KW   Complete proteome.
SQ   SEQUENCE   683 AA;  74518 MW;  F91407FA7094AAD1 CRC64;


  Query Match             29.9%;  Score 59;  DB 2;  Length 683;
  Best Local Similarity   43.8%;  Pred. No. 33;
  Matches   14;  Conservative    2;  Mismatches   12;  Indels    4;  Gaps    1;

Qy       4 PTLRQWLAARAGGGCGGGGIEGPTLRQWLAAR 35
             |  : :||    |       ||||  | || |||
Db      74 PLILEWL----GNAYYRSGIEGAALHQWGAAR 101


RESULT  19
Q9HEA4
ID   Q9HEA4      PRELIMINARY;      PRT;   801 AA.
AC   Q9HEA4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   B11A5.200.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL451109; CAC18624.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_kin_actsite.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   801 AA;  85468 MW;  1BEF5008A0044A33 CRC64;


  Query Match             29.9%;  Score 59;  DB 3;  Length 801;
  Best Local Similarity   53.8%;  Pred. No. 39;
  Matches   14;  Conservative    3;  Mismatches    5;  Indels    4;  Gaps    2;

Qy      15 GGGCGGGGI---EG-PTLRQWLAARA 36
             ||| ||||:   :| |    ||||:|
Db     703 GGGGGGGGVVDDDGEPDFAGWLAAQA 728


RESULT  20
O95692
ID   O95692      PRELIMINARY;   PRT;   805 AA.
AC   O95692;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DJ524E15.1 (PEREGRIN (BR140 PROTEIN)) (FRAGMENT).
GN   DJ524E15.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Palmer S.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84485; CAB06488.1; -.
DR   InterPro; IPR000313; PWWP.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR001965; PHD.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF00855; PWWP; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00249; PHD; 2.
DR   SMART; SM00293; PWWP; 1.
FT   NON_TER       1      1
FT   NON_TER     805    805
SQ   SEQUENCE   805 AA;  90851 MW;  E28C017F5C545334 CRC64;


  Query Match             29.7%;  Score 58.5;  DB 4;  Length 805;
  Best Local Similarity   48.1%;  Pred. No. 45;
  Matches   13;  Conservative    3;  Mismatches    8;  Indels    3;  Gaps    2;

Qy       8 QW-LAARAGGG--CGGGGIEGPTLRQW 31
           ||  |:|| ||  | | |: |   |:|
Db     695 QWGAASRAPGGCSCAGAGLAGGARRRW 721


RESULT  21
Q9JMH4
ID   Q9JMH4      PRELIMINARY;      PRT;  1431 AA.
AC   Q9JMH4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COLLAGEN TYPE XVII.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;
RT   "Mesocricetus auratus mRNA for type XVII collagen.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027759; BAA94381.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 5.
SQ   SEQUENCE   1431 AA;  144579 MW;  4315631FEB2C9A5C CRC64;


  Query Match             29.7%;  Score 58.5;  DB 11;  Length 1431;
  Best Local Similarity   60.0%;  Pred. No. 80;
  Matches   15;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      12 ARAGGGCGGGGIEGPTLRQWLAARA 36
           || ||| ||||  | |   | || |
Db     438 ARGGGGSGGGGGGGGT---WGAAPA 459


RESULT  22
Q9FU26
ID   Q9FU26      PRELIMINARY;      PRT;  117 AA.
AC   Q9FU26;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   P0671B11.11 PROTEIN.
GN   P0671B11.11.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0671B11.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002746; BAB12695.1; -.
SQ   SEQUENCE   117 AA;  12397 MW;  A04617B3DEF9F4B3 CRC64;


  Query Match             29.4%;  Score 58;  DB 10;  Length 117;
  Best Local Similarity   41.9%;  Pred. No. 7.5;
  Matches   13;  Conservative    2;  Mismatches    6;  Indels   10;  Gaps    1;

Qy       8 QWLAARAGGGCGGGG----------IEGPTL 28
           :|| |||  |||  |          : ||||
Db      36 RWLFARATAGCGSQGDDQKKTPEIEVVGPTL 66


RESULT  23
Q56434
ID   Q56434      PRELIMINARY;      PRT;  134 AA.
AC   Q56434;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBOSOMAL PROTEIN L11 (FRAGMENT).
GN   RPL11.
OS   Thermus aquaticus (subsp. thermophilus).
OC   Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94124036; PubMed=8294036;
RA   Heinrich T., Erdmann V.A., Hartmann R.K.;
RT   "Sequence of the gene encoding ribosomal protein L11 from Thermus
RT   thermophilus HB8.";
RL   Gene 136:373-374(1993).
DR   EMBL; L10371; AAA27503.1; -.
KW   Ribosomal protein.
FT   NON_TER       1      1
SQ   SEQUENCE   134 AA;  14336 MW;  C8EB750B7B75EFFC CRC64;


  Query Match             29.4%;  Score 58;  DB 2;  Length 134;
  Best Local Similarity   34.4%;  Pred. No. 8.6;
  Matches   11;  Conservative    4;  Mismatches    5;  Indels   12;  Gaps    1;

Qy      17 GCGGGG----------IEGPTLRQWLAARA 36
           ||||||          : ||: |:|   |:
Db      94 GCGGGGRSGGEGCLSTASVTGPSWRRWTPTRS 125


RESULT  24
Q9W033
ID   Q9W033      PRELIMINARY;      PRT;  170 AA.
AC   Q9W033;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG13807 PROTEIN.
GN   CG13807.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
```

```
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003474; AAF47627.1; -.
DR   FlyBase; FBgn0035323; CG13807.
DR   InterPro; IPR002952; Eggshell.
DR   PRINTS; PR01228; EGGSHELL.
SQ   SEQUENCE   170 AA;  19099 MW;  477D79D55ADF4CE5 CRC64;
```

```
  Query Match             29.4%;  Score 58;  DB 5;  Length 170;
  Best Local Similarity   45.8%;  Pred. No. 11;
  Matches   11;  Conservative    3;  Mismatches    6;  Indels    4;  Gaps    1;

Qy       2 EGPTLRQWLAARAGGGCGGGGIEG 25
           | | :  |:    ||| |||| :|
Db      47 EPPIVENWM----GGGGGGGGFQG 66
```

```
RESULT  25
Q99034
ID   Q99034      PRELIMINARY;      PRT;   302 AA.
AC   Q99034;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ACETYLXYLAN ESTERASE PRECURSOR (EC 3.1.1.72).
GN   AXE1.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 158-186, AND CHARACTERIZATION.
RC   STRAIN=RUTC-30;
RX   MEDLINE=96235218; PubMed=8647098;
RA   Margolles-Clark E., Tenkanen M., Soederlund H., Penttilae M.;
RT   "Acetyl xylan esterase from Trichoderma reesei contains an active-site
RT   serine residue and a cellulose-binding domain.";
RL   Eur. J. Biochem. 237:553-560(1996).
RN   [2]
RP   FUNCTION.
RC   STRAIN=RUTC-30;
RA   Poutanen K., Sundberg M., Korte H., Puls J.;
RT   "Deacetylation of xylans by acetyl esterases of Trichoderma reesei.";
RL   Appl. Microbiol. Biotechnol. 33:506-510(1990).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=RUTC-30;
RA   Sundberg M., Poutanen K.;
RT   "Purification and properties of two acetylxylan esterases of
RT   Trichoderma reesei.";
RL   Biotechnol. Appl. Biochem. 13:1-11(1991).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=98437545; PubMed=9761918;
RA   Hakulinen N., Tenkanen M., Rouvinen J.;
RT   "Crystallization and preliminary X-ray diffraction studies of the
RT   catalytic core of acetyl xylan esterase from Trichoderma reesei.";
RL   Acta Crystallogr. D 54:430-432(1998).
CC   -!- FUNCTION: DEGRADEDS ACETYLATED XYLANS BY CLEAVING ACETYL SIDE
CC       GROUPS FROM THE HETERO-XYLAN BACKBONE.
CC   -!- CATALYTIC ACTIVITY: DEACETYLATION OF XYLANS AND XYLO-
CC       OLIGOSACCHARIDES.
CC   -!- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLOURIDE.
CC   -!- PATHWAY: XYLAN DEGRADATION.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- MASS SPECTROMETRY: MW=21806; METHOD=MALDI.
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
DR   EMBL; Z69256; CAA93247.1; -.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR000734; Lipase.
DR   Pfam; PF00734; CBD_1; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; FALSE_NEG.
DR   PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
KW   Cellulose degradation; Hydrolase; Serine esterase; Glycoprotein;
KW   3D-structure; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     31       POTENTIAL.
FT   CHAIN        32    302       ACETYLXYLAN ESTERASE.
FT   DOMAIN      244    266       LINKER (BY SIMILARITY).
FT   DOMAIN      267    302       CELLULOSE-BINDING (BY SIMILARITY).
FT   MOD_RES      32     32       BLOCKED.
FT   ACT_SITE    121    121       BY SIMILARITY.
FT   DISULFID    274    291       BY SIMILARITY.
FT   DISULFID    285    301       BY SIMILARITY.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (PROBABLE).
SQ   SEQUENCE   302 AA;  30754 MW;  BB6EDCA2971A9F2A CRC64;
```

```
  Query Match             29.4%;  Score 58;  DB 3;  Length 302;
  Best Local Similarity   35.9%;  Pred. No. 19;
  Matches   14;  Conservative    1;  Mismatches    8;  Indels   16;  Gaps    2;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT----------LRQW 31
           |||   |       | ||| |  |||          : ||
Db     265 GPTQTHW------GQCGGQGWTGPTQCESGTTCQVISQW 297
```

```
RESULT  26
Q9XEJ0
```

```
ID   Q9XEJ0      PRELIMINARY;      PRT;   516 AA.
AC   Q9XEJ0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 48.0 KDA PROTEIN.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Llaca V., Lou A., Young S., Messing J.;
RT   "Comparative Analysis of the 22-kDa zein cluster in Z. mays.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF090447; AAD20310.1; -.
DR   InterPro; IPR002952; Eggshell.
DR   PRINTS; PR01228; EGGSHELL.
KW   Hypothetical protein.
SQ   SEQUENCE   516 AA;  48014 MW;  6B40A6043122307A CRC64;


  Query Match             29.4%;  Score 58;  DB 10;  Length 516;
  Best Local Similarity   76.9%;  Pred. No. 33;
  Matches   10;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      15 GGGCGGGGIEGPT 27
           |||||||| | |
Db     359 GGGCGGGGCGGAT 371


RESULT  27
Q9D384
ID   Q9D384      PRELIMINARY;      PRT;   244 AA.
AC   Q9D384;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   6330548G22RIK PROTEIN.
GN   6330548G22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=MEDULLA OBLONGATA;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK018232; BAB31127.1; -.
DR   MGD; MGI:1923417; 6330548G22Rik.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ   SEQUENCE   244 AA;  29290 MW;  1625D74743CE1245 CRC64;


  Query Match             29.2%;  Score 57.5;  DB 11;  Length 244;
  Best Local Similarity   41.9%;  Pred. No. 18;
  Matches   13;  Conservative    2;  Mismatches    7;  Indels    9;  Gaps    1;

Qy       5 TLRQWLAARAGGGCGG---------GGIEGP 26
             ||| |:  | ||| ||          || : |
Db     131 TLRGWIPRRLGGGLGGKKESGQLRFGGRDRP 161


RESULT  28
Q9S596
ID   Q9S596      PRELIMINARY;      PRT;   302 AA.
AC   Q9S596;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PENICILLIN-RESISTANT DD-CARBOXYPEPTIDASE.
GN   PDCA.
OS   Myxococcus xanthus.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX   NCBI_TaxID=34;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99350441; PubMed=10419975;
RA   Kimura Y., Takashima Y., Tokumasu Y., Sato M.;
RT   "Molecular cloning, sequence analysis, and characterization of a
RT   penicillin-resistant DD-carboxypeptidase of Myxococcus xanthus.";
RL   J. Bacteriol. 181:4696-4699(1999).
DR   EMBL; AB023893; BAA83081.1; -.
DR   HSSP; P00733; 1LBU.
DR   InterPro; IPR002477; PG_binding.
DR   InterPro; IPR003709; VanY.
DR   Pfam; PF01471; PG_binding_1; 2.
DR   Pfam; PF02557; VanY; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   302 AA;  31181 MW;  7C844BC85B9E67B7 CRC64;


  Query Match             29.2%;  Score 57.5;  DB 2;  Length 302;
  Best Local Similarity   48.3%;  Pred. No. 22;
  Matches   14;  Conservative    2;  Mismatches    8;  Indels    5;  Gaps    2;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLR 29
           | ||  : | |  : || ||    |||||
Db     75 IVGP--KTWSALNSAGGAGG---SGPTLR 98


RESULT  29
O33230
ID   O33230      PRELIMINARY;      PRT;   495 AA.
AC   O33230;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 53.3 KDA PROTEIN.
GN   HFLX OR RV2725C OR MTCY154.05C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
```

```
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sgares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z98209; CAB10901.1; -.
DR   TubercuList; Rv2725c; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   495 AA;  53327 MW;  F82BA93092945121 CRC64;


  Query Match             29.2%;  Score 57.5;  DB 2;  Length 495;
  Best Local Similarity   43.3%;  Pred. No. 36;
  Matches   13;  Conservative    1;  Mismatches    9;  Indels    7;  Gaps    1;

Qy       4 PTLRQW-------LAARAGGGCGGGGIEGP 26
           | || |         |||| || |: ||
Db     199 PRLRGWGESMSRQAGGRAGGSGGGVGLRGP 228


RESULT  30
Q9C7W8
ID   Q9C7W8      PRELIMINARY;      PRT;    76 AA.
AC   Q9C7W8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 7.9 KDA PROTEIN.
GN   F13N6.12.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
DR   EMBL; AC058785; AAG51509.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   76 AA;  7855 MW;  299A412EA9925CB0 CRC64;


  Query Match             28.9%;  Score 57;  DB 10;  Length 76;
  Best Local Similarity   73.3%;  Pred. No. 6.4;
  Matches   11;  Conservative    1;  Mismatches    1;  Indels    2;  Gaps    1;

Qy       7 RQWLAARAGGGCGGG 21
           |:|||  ||||| ||
Db      64 RRWLA--AGGGCSGG 76


RESULT  31
Q9YHD0
ID   Q9YHD0      PRELIMINARY;      PRT;   377 AA.
AC   Q9YHD0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   OTX.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tomsa J.M., Langeland J.A.;
RT   "Otx expression during lamprey embryogenesis provides insights into
RT   the evolution of the vertebrate head and jaw.";
RL   Dev. Biol. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER HOMEOBOX DOMAINS.
DR   EMBL; AF099746; AAC82470.1; -.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
SQ   SEQUENCE   377 AA;  37998 MW;  C2DBC19402D3A172 CRC64;


  Query Match             28.9%;  Score 57;  DB 13;  Length 377;
  Best Local Similarity   44.4%;  Pred. No. 32;
  Matches   12;  Conservative    2;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       2 EGPTLRQWLAARAGGGCGGGGIEGPTL 28
           :| |   :     ||| ||||  || |
Db     265 QGYTAASYYGVECGGGGGGGGGGPYL 291


RESULT  32
Q9HFM0
ID   Q9HFM0      PRELIMINARY;      PRT;   414 AA.
AC   Q9HFM0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE ENDOCHITINASE CHI2 (FRAGMENT).
GN   CHI2.
OS   Metarhizium anisopliae var. acridum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX   NCBI_TaxID=92637;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FI-985 (ARSEF 324);
RA   Screen S.E., St Leger R.J.;
RT   "Cloning, expression and analysis of chitinase genes from the
RT   entomopathogenic fungus Metarhizium anisopliae.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ293217; CAC07216.1; -.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001579; Chitinase_2.
DR   Pfam; PF00734; CBD_1; 1.
DR   Pfam; PF00192; chitinase_2; 2.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
FT   NON_TER       1       1
SQ   SEQUENCE   414 AA;  43307 MW;  D4CE8945B53CD3AD CRC64;
```

```
  Query Match             28.9%;  Score 57;  DB 3;  Length 414;
  Best Local Similarity   48.0%;  Pred. No. 35;
  Matches   12;  Conservative    1;  Mismatches    6;  Indels    6;  Gaps    1;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPT 27
           | |: ||      | ||| |  |||
Db     377 GGTVPQW------GQCGGEGYSGPT 395


RESULT  33
Q9BZE0
ID   Q9BZE0      PRELIMINARY;      PRT;   524 AA.
AC   Q9BZE0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   KRUPPEL-LIKE ZINC FINGER PROTEIN GLIS2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RA   Zhang F., Kurebayashi S., Jetten A.M.;
RT   "Cloning and genomic structure of GLIS2, a novel gene encoding a Gli-
RT   related, Kruppel-like protein.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF325914; AAK00954.1; -.
SQ   SEQUENCE   524 AA;  55704 MW;  3E2C27243DE5A85E CRC64;


  Query Match             28.9%;  Score 57;  DB 4;  Length 524;
  Best Local Similarity   53.8%;  Pred. No. 44;
  Matches   14;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GPTLRQWLAARAGGGCGGGGIEGPTL 28
           ||     ||   ||| ||||  || |
Db     377 GPLDLSALACGNGGGSGGGGGMGPGL 402


RESULT  34
Q9ASE5
ID   Q9ASE5      PRELIMINARY;      PRT;   529 AA.
AC   Q9ASE5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   P0456F08.14 PROTEIN.
GN   P0456F08.14.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0456F08.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002901; BAB39414.1; -.
SQ   SEQUENCE   529 AA;  55981 MW;  0A5DA55CDD076D24 CRC64;


  Query Match             28.9%;  Score 57;  DB 10;  Length 529;
  Best Local Similarity   63.2%;  Pred. No. 44;
  Matches   12;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 LRQWLAARAGGGCGGGGIE 24
           || : |||: || |||| |
Db     151 LRAYQAARSAGGGGGGGKE 169


RESULT  35
Q9L8D4
ID   Q9L8D4      PRELIMINARY;      PRT;   607 AA.
AC   Q9L8D4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 66.3 KDA PROTEIN (FRAGMENT).
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SO CE90;
RX   MEDLINE=20130945; PubMed=10662695;
RA   Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA   Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA   Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT   "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT   epothilones A and B from Sorangium cellulosum So ce90.";
RL   Chem. Biol. 7:97-109(2000).
DR   EMBL; AF210843; AAF26904.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   607 AA;  66326 MW;  F113CA299B25048E CRC64;


  Query Match             28.9%;  Score 57;  DB 2;  Length 607;
  Best Local Similarity   32.0%;  Pred. No. 51;
  Matches   16;  Conservative    3;  Mismatches    7;  Indels   24;  Gaps    2;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGP----------------TLRQWLAA 34
             ::|| | :|||||         |  ||                  | ||||
Db      96 VDGPALVRWLAAR--------GAPGPLREYEEERERARTAQEARRLWLAA 137


RESULT  36
Q9P270
ID   Q9P270      PRELIMINARY;      PRT;   612 AA.
AC   Q9P270;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   KIAA1458 PROTEIN (FRAGMENT).
GN   KIAA1458.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human
RT   genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT   which code for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
DR   EMBL; AB040891; BAA95982.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   612 AA;  65593 MW;  9AA4061D21E1E9FD CRC64;


  Query Match             28.9%;  Score 57;  DB 4;  Length 612;
  Best Local Similarity   59.1%;  Pred. No. 51;
  Matches   13;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 PTLRQWLAARAGGGCGGGGIEG 25
           |:|   |  ||||| ||||  |
Db      10 PSLSLSLRERAGGGGGGGGGAG 31
```

```
RESULT  37
Q9LGW5
ID   Q9LGW5      PRELIMINARY;      PRT;   651 AA.
AC   Q9LGW5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0706B05.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002482; BAA96618.1; -.
SQ   SEQUENCE   651 AA;  69800 MW;  0308FB36B83B62B0 CRC64;


  Query Match             28.9%;  Score 57;  DB 10;  Length 651;
  Best Local Similarity   61.1%;  Pred. No. 55;
  Matches   11;  Conservative    2;  Mismatches    1;  Indels    4;  Gaps    1;

Qy      12 ARAGGG----CGGGGIEG 25
           | :|||    |||||:||
Db     418 AASGGGFFCTCGGGGVEG 435


RESULT  38
O75182
ID   O75182      PRELIMINARY;      PRT;  1130 AA.
AC   O75182;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   KIAA0700 PROTEIN (FRAGMENT).
GN   KIAA0700.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
DR   EMBL; AB014600; BAA31675.1; -.
DR   InterPro; IPR003822; PAH.
DR   Pfam; PF02671; PAH; 3.
FT   NON_TER       1       1
SQ   SEQUENCE   1130 AA;  129358 MW;  B767339317ECC96D CRC64;


  Query Match             28.9%;  Score 57;  DB 4;  Length 1130;
  Best Local Similarity   54.2%;  Pred. No. 94;
  Matches   13;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      12 ARAGGGCGGGGIEGPTLRQWLAAR 35
           | |||| || |  ||  |    ||
Db       2 AHAGGGSGGGSGAGGPAGRGLSGAR 25


RESULT  39
Q9YDB1
ID   Q9YDB1      PRELIMINARY;      PRT;   176 AA.
AC   Q9YDB1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 19.2 KDA PROTEIN APE1002.
GN   APE1002.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000060; BAA79986.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   176 AA;  19234 MW;  684D7F476A254457 CRC64;


  Query Match             28.7%;  Score 56.5;  DB 1;  Length 176;
  Best Local Similarity   34.9%;  Pred. No. 17;
  Matches   15;  Conservative    1;  Mismatches    8;  Indels   19;  Gaps    1;

Qy       7 RQWLAARAGGGC------------------GGGGIEGPTLRQ 30
           || |    ||||                  |||| ||  ||:
Db      12 RQGLHGEEGGGCDCKGCGRRLNPPPPGHHWQGGGGEEGEELRR 54


RESULT  40
Q9AR44
ID   Q9AR44      PRELIMINARY;      PRT;   243 AA.
AC   Q9AR44;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   P0498A12.7 PROTEIN (OSJNBA0004B13.18 PROTEIN).
GN   P0498A12.7.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0498A12.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OSJNBa0004B13.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003020; BAB39979.1; -.
DR   EMBL; AP003018; BAB39964.1; -.
SQ   SEQUENCE   243 AA;  26243 MW;  029E9344C20E0EC8 CRC64;
```

```
  Query Match             28.7%;  Score 56.5;  DB 10;  Length 243;
  Best Local Similarity   47.8%;  Pred. No. 23;
  Matches   11;  Conservative    3;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      1 IEGPTLRQWLAARAGGGCGGGGI 23
          :  |  |:|   | |||  ||||:
Db     28 VRTPDQRRW---RRGGGGGGGGV 47


RESULT  41
Q9NYW9
ID   Q9NYW9      PRELIMINARY;      PRT;  1548 AA.
AC   Q9NYW9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SOMATOSTATIN RECEPTOR-INTERACTING PROTEIN SPLICE VARIANT B.
GN   SSTRIP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20020275; PubMed=10551867;
RA   Zitzer H., Honck H.H., Bachner D., Richter D., Kreienkamp H.J.;
RT   "Somatostatin receptor interacting protein defines a novel family of
RT   multidomain proteins present in human and rodent brain.";
RL   J. Biol. Chem. 274:32997-33001(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Zitzer H., Hoenck H., Baechner D., Richter D., Kreienkamp H.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF226728; AAF35887.1; -.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001660; SAM.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00454; SAM; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Receptor.
SQ   SEQUENCE   1548 AA;  158853 MW;  8E5209F785A04D73 CRC64;


  Query Match             28.7%;  Score 56.5;  DB 4;  Length 1548;
  Best Local Similarity   35.7%;  Pred. No. 1.5e+02;
  Matches   15;  Conservative    2;  Mismatches    6;  Indels   19;  Gaps    2;

Qy      4 PTLRQWLAARAGG---------------GCGGGGIEGPTLR 29
          |:|| |   | ||               | |||  :|| ||
Db    432 PSLRGW---RGGGPSPTPGAPSPSHHGSAGGGGGSSQGPALR 470


RESULT  42
Q9Y566
ID   Q9Y566      PRELIMINARY;      PRT;  2161 AA.
AC   Q9Y566;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SOMATOSTATIN RECEPTOR INTERACTING PROTEIN SPLICE VARIANT A.
GN   SSTRIP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zitzer H., Hoenck H.-H., Richter D., Kreienkamp H.-J.;
RT   "The somatostatin receptor interacting protein (SSTRIP) defines a
RT   novel family of multidomain postsynaptic density proteins.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF163302; AAD45121.1; -.
DR   HSSP; P06241; 1SHF.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001660; SAM.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00248; ANK; 3.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00454; SAM; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50106; PDZ; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Receptor.
SQ   SEQUENCE   2161 AA;  225019 MW;  5FEFC969CBE98701 CRC64;


  Query Match             28.7%;  Score 56.5;  DB 4;  Length 2161;
  Best Local Similarity   35.7%;  Pred. No. 2.1e+02;
  Matches   15;  Conservative    2;  Mismatches    6;  Indels   19;  Gaps    2;

Qy      4 PTLRQWLAARAGG---------------GCGGGGIEGPTLR 29
          |:|| |   | ||               | |||  :|| ||
Db   1045 PSLRGW---RGGGPSPTPGAPSPSHHGSAGGGGGSSQGPALR 1083


RESULT  43
O34781
ID   O34781      PRELIMINARY;      PRT;    56 AA.
AC   O34781;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SUBLANCIN 168 PRECURSOR PEPTIDE.
GN   SUNA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Paik S.H., Hansen J.N.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
```

```
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF014938; AAC63531.1; -.
DR   EMBL; Z99115; CAB14066.1; -.
KW   Complete proteome.
SQ   SEQUENCE   56 AA;  5982 MW;  79EC0BF822F9F4C0 CRC64;


  Query Match             28.4%;  Score 56;  DB 2;  Length 56;
  Best Local Similarity   44.0%;  Pred. No. 6.2;
  Matches   11;  Conservative    3;  Mismatches    9;  Indels    2;  Gaps    1;

Qy       9 WLAARAGG--GCGGGGIEGPTLRQW 31
           ||   :||  ||||| :     ||:
Db      30 WLQCASGGTIGCGGGAVACQNYRQF 54


RESULT  44
O64033
ID   O64033      PRELIMINARY;      PRT;    56 AA.
AC   O64033;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PUTATIVE LIPOPROTEIN.
GN   YOLG.
OS   Bacteriophage SPBc2.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX   NCBI_TaxID=66797;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,
RA   Karamata D.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF020713; AAC12992.1; -.
KW   Lipoprotein.
SQ   SEQUENCE   56 AA;  5982 MW;  79EC0BF822F9F4C0 CRC64;


  Query Match             28.4%;  Score 56;  DB 9;  Length 56;
  Best Local Similarity   44.0%;  Pred. No. 6.2;
  Matches   11;  Conservative    3;  Mismatches    9;  Indels    2;  Gaps    1;

Qy       9 WLAARAGG--GCGGGGIEGPTLRQW 31
           ||   :||  ||||| :     ||:
Db      30 WLQCASGGTIGCGGGAVACQNYRQF 54


RESULT  45
O61832
ID   O61832      PRELIMINARY;      PRT;   163 AA.
AC   O61832;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   C23H5.9 PROTEIN.
GN   C23H5.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Lamar E., Kramer J.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF067609; AAC17537.1; -.
SQ   SEQUENCE   163 AA;  16317 MW;  A068D74244200258 CRC64;


  Query Match             28.4%;  Score 56;  DB 5;  Length 163;
  Best Local Similarity   75.0%;  Pred. No. 18;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    1;

Qy      15 GGGCGGGGIEG--PTL 28
           |||||||  |  |||
Db      33 GGGCGGGGGGGCLPTL 48


Search completed: December 26, 2001, 10:29:14
Job time: 111 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:26:03 ; Search time 12.63 Seconds
(without alignments)
64.142 Million cell updates/sec

Title: US-09-422-838C-33
Perfect score: 197
Sequence: 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	76.5	38.8	25	2	US-08-764-640-231	Sequence 231, App
2	76.5	38.8	25	3	US-09-244-298A-231	Sequence 231, App
3	76.5	38.8	25	4	US-09-516-704-231	Sequence 231, App
4	73	37.1	14	2	US-08-764-640-13	Sequence 13, Appl
5	73	37.1	14	2	US-08-764-640-193	Sequence 193, App
6	73	37.1	14	3	US-08-973-225-13	Sequence 13, Appl
7	73	37.1	14	3	US-08-973-225-193	Sequence 193, App
8	73	37.1	14	3	US-09-244-298A-13	Sequence 13, Appl
9	73	37.1	14	3	US-09-244-298A-193	Sequence 193, App
10	73	37.1	14	4	US-09-516-704-13	Sequence 13, Appl
11	73	37.1	14	4	US-09-516-704-193	Sequence 193, App
12	73	37.1	15	2	US-08-764-640-17	Sequence 17, Appl
13	73	37.1	15	2	US-08-764-640-185	Sequence 185, App
14	73	37.1	15	3	US-08-973-225-17	Sequence 17, Appl
15	73	37.1	15	3	US-08-973-225-185	Sequence 185, App
16	73	37.1	15	3	US-09-244-298A-17	Sequence 17, Appl
17	73	37.1	15	3	US-09-244-298A-185	Sequence 185, App
18	73	37.1	15	4	US-09-516-704-17	Sequence 17, Appl
19	73	37.1	15	4	US-09-516-704-185	Sequence 185, App
20	73	37.1	16	2	US-08-764-640-18	Sequence 18, Appl
21	73	37.1	16	2	US-08-764-640-194	Sequence 194, App
22	73	37.1	16	2	US-08-764-640-232	Sequence 232, App
23	73	37.1	16	3	US-08-973-225-18	Sequence 18, Appl
24	73	37.1	16	3	US-08-973-225-194	Sequence 194, App
25	73	37.1	16	3	US-08-973-225-220	Sequence 220, App
26	73	37.1	16	3	US-09-244-298A-18	Sequence 18, Appl
27	73	37.1	16	3	US-09-244-298A-194	Sequence 194, App
28	73	37.1	16	3	US-09-244-298A-232	Sequence 232, App
29	73	37.1	16	4	US-09-516-704-18	Sequence 18, Appl
30	73	37.1	16	4	US-09-516-704-194	Sequence 194, App
31	73	37.1	16	4	US-09-516-704-232	Sequence 232, App
32	69	35.0	14	2	US-08-764-640-195	Sequence 195, App
33	69	35.0	14	2	US-08-764-640-199	Sequence 199, App
34	69	35.0	14	3	US-08-973-225-195	Sequence 195, App
35	69	35.0	14	3	US-08-973-225-199	Sequence 199, App
36	69	35.0	14	3	US-09-244-298A-195	Sequence 195, App
37	69	35.0	14	3	US-09-244-298A-199	Sequence 199, App
38	69	35.0	14	4	US-09-516-704-195	Sequence 195, App
39	69	35.0	14	4	US-09-516-704-199	Sequence 199, App
40	69	35.0	15	2	US-08-764-640-196	Sequence 196, App
41	69	35.0	15	2	US-08-764-640-200	Sequence 200, App
42	69	35.0	15	2	US-08-764-640-209	Sequence 209, App
43	69	35.0	15	2	US-08-764-640-215	Sequence 215, App
44	69	35.0	15	3	US-08-973-225-196	Sequence 196, App
45	69	35.0	15	3	US-08-973-225-200	Sequence 200, App
46	69	35.0	15	3	US-09-244-298A-196	Sequence 196, App
47	69	35.0	15	3	US-09-244-298A-200	Sequence 200, App
48	69	35.0	15	3	US-09-244-298A-209	Sequence 209, App
49	69	35.0	15	3	US-09-244-298A-215	Sequence 215, App
50	69	35.0	15	4	US-09-516-704-196	Sequence 196, App
51	69	35.0	15	4	US-09-516-704-200	Sequence 200, App
52	69	35.0	15	4	US-09-516-704-209	Sequence 209, App
53	69	35.0	15	4	US-09-516-704-215	Sequence 215, App
54	69	35.0	16	2	US-08-764-640-210	Sequence 210, App
55	69	35.0	16	3	US-09-244-298A-210	Sequence 210, App
56	69	35.0	16	4	US-09-516-704-210	Sequence 210, App
57	67	34.0	15	2	US-08-764-640-211	Sequence 211, App
58	67	34.0	15	2	US-08-764-640-212	Sequence 212, App
59	67	34.0	15	3	US-09-244-298A-211	Sequence 211, App
60	67	34.0	15	3	US-09-244-298A-212	Sequence 212, App
61	67	34.0	15	4	US-09-516-704-211	Sequence 211, App
62	67	34.0	15	4	US-09-516-704-212	Sequence 212, App
63	66	33.5	13	2	US-08-764-640-197	Sequence 197, App
64	66	33.5	13	3	US-08-973-225-197	Sequence 197, App
65	66	33.5	13	3	US-09-244-298A-197	Sequence 197, App
66	66	33.5	13	4	US-09-516-704-197	Sequence 197, App
67	65	33.0	13	2	US-08-764-640-201	Sequence 201, App
68	65	33.0	13	3	US-08-973-225-201	Sequence 201, App
69	65	33.0	13	3	US-09-244-298A-201	Sequence 201, App
70	65	33.0	13	4	US-09-516-704-201	Sequence 201, App
71	65	33.0	14	2	US-08-764-640-202	Sequence 202, App
72	65	33.0	14	3	US-08-973-225-202	Sequence 202, App
73	65	33.0	14	3	US-09-244-298A-202	Sequence 202, App
74	65	33.0	14	4	US-09-516-704-202	Sequence 202, App
75	64	32.5	14	2	US-08-764-640-198	Sequence 198, App
76	64	32.5	14	3	US-08-973-225-198	Sequence 198, App
77	64	32.5	14	3	US-09-244-298A-198	Sequence 198, App
78	64	32.5	14	4	US-09-516-704-198	Sequence 198, App
79	62.5	31.7	130	1	US-08-451-947-17	Sequence 17, Appl
80	62.5	31.7	130	1	US-08-451-947-22	Sequence 22, Appl
81	62.5	31.7	130	2	US-08-424-826A-17	Sequence 17, Appl
82	62.5	31.7	130	2	US-08-424-826A-22	Sequence 22, Appl
83	62.5	31.7	130	3	US-08-928-694-17	Sequence 17, Appl
84	62.5	31.7	130	3	US-08-928-694-22	Sequence 22, Appl
85	62.5	31.7	130	5	PCT-US91-06950-17	Sequence 17, Appl
86	62.5	31.7	130	5	PCT-US91-06950-22	Sequence 22, Appl
87	62	31.5	12	2	US-08-764-640-203	Sequence 203, App
88	62	31.5	12	3	US-08-973-225-203	Sequence 203, App
89	62	31.5	12	3	US-09-244-298A-203	Sequence 203, App
90	62	31.5	12	4	US-09-516-704-203	Sequence 203, App
91	61.5	31.2	130	1	US-08-451-947-23	Sequence 23, Appl
92	61.5	31.2	130	2	US-08-424-826A-23	Sequence 23, Appl
93	61.5	31.2	130	3	US-08-928-694-23	Sequence 23, Appl
94	61.5	31.2	130	5	PCT-US91-06950-23	Sequence 23, Appl
95	61	31.0	14	2	US-08-764-640-226	Sequence 226, App
96	61	31.0	14	3	US-09-244-298A-226	Sequence 226, App
97	61	31.0	14	4	US-09-516-704-226	Sequence 226, App
98	61	31.0	15	2	US-08-764-640-213	Sequence 213, App
99	61	31.0	15	2	US-08-764-640-227	Sequence 227, App
100	61	31.0	15	3	US-09-244-298A-213	Sequence 213, App

ALIGNMENTS

```
RESULT   1
US-08-764-640-231
; Sequence 231, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  231:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  13
;      OTHER INFORMATION:  /product= "Ava"
US-08-764-640-231


  Query Match             38.8%;  Score 76.5;  DB 2;  Length 25;
  Best Local Similarity   40.6%;  Pred. No. 0.009;
  Matches   13;  Conservative    8;  Mismatches    2;  Indels    9;  Gaps    1;

Qy       2 EGPTLRQWLAARAGGGCGGGGIEGPTLRQWLA 33
           :|||||:|::  |         :|||||:|::
Db       2 DGPTLREWISFXA---------DGPTLREWIS 24


RESULT   2
US-09-244-298A-231
; Sequence 231, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  231:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  13
;      OTHER INFORMATION:  /product= "Ava"
US-09-244-298A-231


  Query Match             38.8%;  Score 76.5;  DB 3;  Length 25;
  Best Local Similarity   40.6%;  Pred. No. 0.009;
  Matches   13;  Conservative    8;  Mismatches    2;  Indels    9;  Gaps    1;

Qy       2 EGPTLRQWLAARAGGGCGGGGIEGPTLRQWLA 33
           :|||||:|::  |         :|||||:|::
Db       2 DGPTLREWISFXA---------DGPTLREWIS 24


RESULT   3
US-09-516-704-231
; Sequence 231, Application US/09516704
; Patent No. 6251864
;  GENERAL INFORMATION:
;       APPLICANT: Dower, William J.
;                  Barrett, Ronald W.
;                  Cwirla, Steven E.
;                  Gates, Christian
;                  Schatz, Peter J.
;                  Balasubramanian, Palaniappan
;                  Wagstrom, Christopher R.
```

```
;                  Hendren, Richard W.
;                  Deprince, Randolph B.
;                  Podduturi, Surekha
;       TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                           RECEPTOR
;       NUMBER OF SEQUENCES: 244
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Glaxo Wellcome
;            STREET: Five Moore Drive, P.O. Box 13398
;            CITY: Research Triangle Park
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/516,704
;            FILING DATE: 01-Mar-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3281
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;  INFORMATION FOR SEQ ID NO: 231:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 25 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       FEATURE:
;            NAME/KEY:  Modified-site
;            LOCATION:  13
;            OTHER INFORMATION:  /product= "Ava"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 231:
US-09-516-704-231


  Query Match             38.8%;  Score 76.5;  DB 4;  Length 25;
  Best Local Similarity   40.6%;  Pred. No. 0.009;
  Matches   13;  Conservative    8;  Mismatches    2;  Indels    9;  Gaps    1;

Qy       2 EGPTLRQWLAARAGGGCGGGGIEGPTLRQWLA 33
           :|||||:|::  |         :||||||:|::
Db       2 DGPTLREWISFXA---------DGPTLREWIS 24


RESULT   4
US-08-764-640-13
; Sequence 13, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-13


  Query Match             37.1%;  Score 73;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARA 14
           ||||||||||||||
Db       1 IEGPTLRQWLAARA 14


RESULT   5
US-08-764-640-193
; Sequence 193, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  193:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-193


  Query Match             37.1%;  Score 73;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT   6
US-08-973-225-13
; Sequence 13, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 13:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13


  Query Match             37.1%;  Score 73;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT   7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 193:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-08-973-225-193


  Query Match             37.1%;  Score 73;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14
```

```
RESULT   8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-13


  Query Match             37.1%;  Score 73;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT   9
US-09-244-298A-193
; Sequence 193, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  193:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-193


  Query Match             37.1%;  Score 73;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  10
US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
```

```
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 13:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13


  Query Match             37.1%;  Score 73;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  11
US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 193:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193


  Query Match             37.1%;  Score 73;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  12
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-17
```

```
  Query Match             37.1%;  Score 73;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  13
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  185:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-185


  Query Match             37.1%;  Score 73;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  14
US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17


  Query Match             37.1%;  Score 73;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  15
US-08-973-225-185
; Sequence 185, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
```

```
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 185:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185


  Query Match             37.1%;  Score 73;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  16
US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-17


  Query Match             37.1%;  Score 73;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  17
US-09-244-298A-185
; Sequence 185, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  185:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-185


  Query Match             37.1%;  Score 73;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  18
US-09-516-704-17
; Sequence 17, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17


  Query Match             37.1%;  Score 73;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  19
US-09-516-704-185
; Sequence 185, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 185:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-09-516-704-185


  Query Match             37.1%;  Score 73;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  20
US-08-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
```

```
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  15
;      OTHER INFORMATION:  /product= "Beta-ala"
US-08-764-640-18


  Query Match             37.1%;  Score 73;  DB 2;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  21
US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-194


  Query Match             37.1%;  Score 73;  DB 2;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  22
US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  232:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-232


  Query Match             37.1%;  Score 73;  DB 2;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  23
US-08-973-225-18
; Sequence 18, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 18:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        FEATURE:
;             NAME/KEY:  Modified-site
;             LOCATION:  15
;             OTHER INFORMATION:  /product= "Beta-ala"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 194:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
```

```
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          |||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
;  GENERAL INFORMATION:
;       APPLICANT: Dower, William J.
;                  Barrett, Ronald W.
;                  Cwirla, Steven E.
;                  Duffin, David J.
;                  Gates, Christian
;                  Haselden, Sherril S.
;                  Mattheakis, Larry C.
;                  Schatz, Peter J.
;                  Wagstrom, Christopher R.
;                  Wrighton, Nicholas C.
;       TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                           THROMBOPOIETIN RECEPTOR
;       NUMBER OF SEQUENCES: 232
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Glaxo Wellcome
;            STREET: Five Moore Drive, P.O. Box 13398
;            CITY: Research Triangle Park
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/973,225A
;            FILING DATE: 04-Dec-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3065USW
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;  INFORMATION FOR SEQ ID NO: 220:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 16 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          |||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  18:-
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  15
;      OTHER INFORMATION:  /product= "Beta-ala"
US-09-244-298A-18


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          |||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  27
US-09-244-298A-194
; Sequence 194, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
```

```
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-194


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARA 14
           ||||||||||||||
Db       2 IEGPTLRQWLAARA 15


RESULT  28
US-09-244-298A-232
; Sequence 232, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  232:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-232


  Query Match             37.1%;  Score 73;  DB 3;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARA 14
           ||||||||||||||
Db       2 IEGPTLRQWLAARA 15


RESULT  29
US-09-516-704-18
; Sequence 18, Application US/09516704
; Patent No. 6251864
;  GENERAL INFORMATION:
;       APPLICANT: Dower, William J.
;                  Barrett, Ronald W.
;                  Cwirla, Steven E.
;                  Gates, Christian
;                  Schatz, Peter J.
;                  Balasubramanian, Palaniappan
;                  Wagstrom, Christopher R.
;                  Hendren, Richard W.
;                  Deprince, Randolph B.
;                  Podduturi, Surekha
;       TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                           RECEPTOR
;       NUMBER OF SEQUENCES: 244
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Glaxo Wellcome
;            STREET: Five Moore Drive, P.O. Box 13398
;            CITY: Research Triangle Park
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/516,704
;            FILING DATE: 01-Mar-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3281
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;  INFORMATION FOR SEQ ID NO: 18:
;       SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         FEATURE:
;             NAME/KEY:  Modified-site
;             LOCATION:  15
;             OTHER INFORMATION:  /product= "Beta-ala"
;         SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-516-704-18


  Query Match             37.1%;  Score 73;  DB 4;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      1 IEGPTLRQWLAARA 14


RESULT  30
US-09-516-704-194
; Sequence 194, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;         APPLICANT: Dower, William J.
;                    Barrett, Ronald W.
;                    Cwirla, Steven E.
;                    Gates, Christian
;                    Schatz, Peter J.
;                    Balasubramanian, Palaniappan
;                    Wagstrom, Christopher R.
;                    Hendren, Richard W.
;                    Deprince, Randolph B.
;                    Podduturi, Surekha
;         TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                             RECEPTOR
;         NUMBER OF SEQUENCES: 244
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;    INFORMATION FOR SEQ ID NO: 194:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-09-516-704-194


  Query Match             37.1%;  Score 73;  DB 4;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  31
US-09-516-704-232
; Sequence 232, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;         APPLICANT: Dower, William J.
;                    Barrett, Ronald W.
;                    Cwirla, Steven E.
;                    Gates, Christian
;                    Schatz, Peter J.
;                    Balasubramanian, Palaniappan
;                    Wagstrom, Christopher R.
;                    Hendren, Richard W.
;                    Deprince, Randolph B.
;                    Podduturi, Surekha
;         TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                             RECEPTOR
;         NUMBER OF SEQUENCES: 244
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;    INFORMATION FOR SEQ ID NO: 232:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         SEQUENCE DESCRIPTION: SEQ ID NO: 232:
US-09-516-704-232


  Query Match             37.1%;  Score 73;  DB 4;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARA 14
          ||||||||||||||
Db      2 IEGPTLRQWLAARA 15


RESULT  32
US-08-764-640-195
; Sequence 195, Application US/08764640
; Patent No. 5869451
```

```
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  195:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-195


  Query Match             35.0%;  Score 69;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAAR 13
           |||||||||||||
Db       1 IEGPTLRQWLAAR 13


RESULT  33
US-08-764-640-199
; Sequence 199, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  199:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-764-640-199


  Query Match             35.0%;  Score 69;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EGPTLRQWLAARA 14
           |||||||||||||
Db       1 EGPTLRQWLAARA 13


RESULT  34
US-08-973-225-195
; Sequence 195, Application US/08973225A
; Patent No. 6083913
;  GENERAL INFORMATION:
;       APPLICANT:  Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;       TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;       NUMBER OF SEQUENCES:  232
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE:  Glaxo Wellcome
;            STREET:  Five Moore Drive, P.O. Box 13398
;            CITY:  Research Triangle Park
;            STATE:  NC
;            COUNTRY:  USA
;            ZIP:  27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE:  Floppy disk
;            COMPUTER:  IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 195:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 195:
US-08-973-225-195


  Query Match             35.0%;  Score 69;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAAR 13
          |||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  35
US-08-973-225-199
; Sequence 199, Application US/08973225A
; Patent No. 6083913
;  GENERAL INFORMATION:
;       APPLICANT: Dower, William J.
;                  Barrett, Ronald W.
;                  Cwirla, Steven E.
;                  Duffin, David J.
;                  Gates, Christian
;                  Haselden, Sherril S.
;                  Mattheakis, Larry C.
;                  Schatz, Peter J.
;                  Wagstrom, Christopher R.
;                  Wrighton, Nicholas C.
;       TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                           THROMBOPOIETIN RECEPTOR
;       NUMBER OF SEQUENCES: 232
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Glaxo Wellcome
;            STREET: Five Moore Drive, P.O. Box 13398
;            CITY: Research Triangle Park
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/973,225A
;            FILING DATE: 04-Dec-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3065USW
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 199:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 14 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-08-973-225-199


  Query Match             35.0%;  Score 69;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EGPTLRQWLAARA 14
          |||||||||||||
Db      1 EGPTLRQWLAARA 13


RESULT  36
US-09-244-298A-195
; Sequence 195, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  195:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-195


  Query Match             35.0%;  Score 69;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAAR 13
```

```
          ||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  37
US-09-244-298A-199
; Sequence 199, Application US/09244298A
; Patent No. 6121238
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,298A
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  199:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-244-298A-199


  Query Match             35.0%;  Score 69;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EGPTLRQWLAARA 14
          |||||||||||||
Db      1 EGPTLRQWLAARA 13


RESULT  38
US-09-516-704-195
; Sequence 195, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/516,704
;             FILING DATE: 01-Mar-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3281
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 195:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 195:
US-09-516-704-195


  Query Match             35.0%;  Score 69;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAAR 13
          |||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  39
US-09-516-704-199
; Sequence 199, Application US/09516704
; Patent No. 6251864
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Gates, Christian
;                   Schatz, Peter J.
;                   Balasubramanian, Palaniappan
;                   Wagstrom, Christopher R.
;                   Hendren, Richard W.
;                   Deprince, Randolph B.
;                   Podduturi, Surekha
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            RECEPTOR
;        NUMBER OF SEQUENCES: 244
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
```

```
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/516,704
;            FILING DATE: 01-Mar-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3281
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;  INFORMATION FOR SEQ ID NO: 199:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 14 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-09-516-704-199


  Query Match            35.0%;  Score 69;  DB 4;  Length 14;
  Best Local Similarity  100.0%;  Pred. No. 0.036;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EGPTLRQWLAARA 14
          |||||||||||||
Db      1 EGPTLRQWLAARA 13


RESULT  40
US-08-764-640-196
; Sequence 196, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  196:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  14
;      OTHER INFORMATION:  /product= "Beta-ala"
US-08-764-640-196


  Query Match            35.0%;  Score 69;  DB 2;  Length 15;
  Best Local Similarity  100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAAR 13
          |||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  41
US-08-764-640-200
; Sequence 200, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  200:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  14
;      OTHER INFORMATION:  /product= "Beta-ala"
US-08-764-640-200


  Query Match             35.0%;  Score 69;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EGPTLRQWLAARA 14
           |||||||||||||
Db       1 EGPTLRQWLAARA 13


RESULT  42
US-08-764-640-209
; Sequence 209, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  14
;      OTHER INFORMATION:  /product= "N-methyl-Ala"
US-08-764-640-209


  Query Match             35.0%;  Score 69;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAAR 13
           |||||||||||||
Db       1 IEGPTLRQWLAAR 13


RESULT  43
US-08-764-640-215
; Sequence 215, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
;  GENERAL INFORMATION:
;    APPLICANT:  Dower, William J.
;    APPLICANT:  Barrett, Ronald W.
;    APPLICANT:  Cwirla, Steven E.
;    APPLICANT:  Gates, Christian
;    APPLICANT:  Schatz, Peter J.
;    APPLICANT:  Balasubramanian, Palaniappan
;    APPLICANT:  Wagstrom, Christopher R.
;    APPLICANT:  Hendren, Richard W.
;    APPLICANT:  Deprince, Randolph B.
;    APPLICANT:  Podduturi, Surekha
;    APPLICANT:  Yin, Qun
;    TITLE OF INVENTION:  PEPTIDES AND COMPOUNDS THAT BIND TO A
;    TITLE OF INVENTION:  RECEPTOR
;    NUMBER OF SEQUENCES:  244
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Glaxo Wellcome
;      STREET:  Five Moore Drive, P.O. Box 13398
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,640
;      FILING DATE:  11-DEC-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hrubiec, Robert T.
;      REGISTRATION NUMBER:  36,392
;      REFERENCE/DOCKET NUMBER:  PK3281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-248-1000
;  INFORMATION FOR SEQ ID NO:  215:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  14
;      OTHER INFORMATION:  /product= "Sar"
US-08-764-640-215


  Query Match             35.0%;  Score 69;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 IEGPTLRQWLAAR 13
          ||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  44
US-08-973-225-196
; Sequence 196, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 196:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        FEATURE:
;             NAME/KEY:  Modified-site
;             LOCATION:  14
;             OTHER INFORMATION:  /product= "Beta-ala"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 196:
US-08-973-225-196


  Query Match             35.0%;  Score 69;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAAR 13
          ||||||||||||
Db      1 IEGPTLRQWLAAR 13


RESULT  45
US-08-973-225-200
; Sequence 200, Application US/08973225A
; Patent No. 6083913
;   GENERAL INFORMATION:
;        APPLICANT: Dower, William J.
;                   Barrett, Ronald W.
;                   Cwirla, Steven E.
;                   Duffin, David J.
;                   Gates, Christian
;                   Haselden, Sherril S.
;                   Mattheakis, Larry C.
;                   Schatz, Peter J.
;                   Wagstrom, Christopher R.
;                   Wrighton, Nicholas C.
;        TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;                            THROMBOPOIETIN RECEPTOR
;        NUMBER OF SEQUENCES: 232
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Glaxo Wellcome
;             STREET: Five Moore Drive, P.O. Box 13398
;             CITY: Research Triangle Park
;             STATE: NC
;             COUNTRY: USA
;             ZIP: 27709
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/973,225A
;             FILING DATE: 04-Dec-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Hrubiec, Robert T.
;             REGISTRATION NUMBER: 36,392
;             REFERENCE/DOCKET NUMBER: PK3065USW
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 919-248-1000
;   INFORMATION FOR SEQ ID NO: 200:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        FEATURE:
;             NAME/KEY:  Modified-site
;             LOCATION:  14
;             OTHER INFORMATION:  /product= "Beta-ala"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-08-973-225-200


  Query Match             35.0%;  Score 69;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EGPTLRQWLAARA 14
          |||||||||||||
Db      1 EGPTLRQWLAARA 13


Search completed: December 26, 2001, 10:28:21
Job time: 138 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:25:08 ; Search time 23.99 Seconds
(without alignments)
111.156 Million cell updates/sec

Title: US-09-422-838C-33
Perfect score: 197
Sequence: 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	197	100.0	36	21	AAB17303	TPO-mimetic peptid
2	197	100.0	36	21	AAB17307	TPO-mimetic peptid
3	197	100.0	36	21	AAY96524	Thrombopoietin mim
4	185.5	94.2	39	21	AAB17305	TPO-mimetic peptid
5	185	93.9	36	21	AAB16963	TPO-mimetic peptid
6	185	93.9	36	21	AAB17293	TPO-mimetic peptid
7	185	93.9	36	21	AAB17301	TPO-mimetic peptid
8	185	93.9	36	21	AAY96523	Thrombopoietin mim
9	185	93.9	36	21	AAY96525	Thrombopoietin mim
10	185	93.9	40	21	AAB17302	TPO-mimetic peptid
11	185	93.9	41	21	AAY96528	Thrombopoietin mim
12	185	93.9	42	21	AAB17281	TPO-mimetic peptid
13	185	93.9	42	21	AAB17282	TPO-mimetic peptid
14	185	93.9	42	21	AAB17308	Synthetic TMP-TMP
15	185	93.9	42	21	AAY96530	Thrombopoietin mim
16	185	93.9	60	21	AAB17311	Synthetic TMP-TMP-
17	185	93.9	269	21	AAB16960	TMP-TMP-Fc protein
18	185	93.9	269	21	AAY96531	Human IgG1 Fc TMP
19	181	91.9	268	21	AAB16959	Fc-TMP-TMP protein
20	179	90.9	36	21	AAB17306	TPO-mimetic peptid
21	179	90.9	36	21	AAY96526	Thrombopoietin mim
22	177.5	90.1	35	21	AAB17292	TPO-mimetic peptid
23	174.5	88.6	37	21	AAB17294	TPO-mimetic peptid
24	174	88.3	38	21	AAB17295	TPO-mimetic peptid
25	173.5	88.1	39	21	AAB17304	TPO-mimetic peptid
26	172	87.3	42	21	AAB17296	TPO-mimetic peptid
27	171	86.8	34	21	AAB17291	TPO-mimetic peptid
28	164.5	83.5	33	21	AAB17290	TPO-mimetic peptid
29	159	80.7	36	21	AAB17298	TPO-mimetic peptid
30	159	80.7	36	21	AAB17299	TPO-mimetic peptid
31	159	80.7	36	21	AAY96521	Cyclic or linear t
32	158	80.2	32	21	AAB17289	TPO-mimetic peptid
33	157	79.7	36	21	AAB17300	TPO-mimetic peptid
34	157	79.7	36	21	AAY96522	Linear thrombopoie
35	151.5	76.9	31	21	AAB17288	TPO-mimetic peptid
36	145	73.6	30	21	AAB17287	TPO-mimetic peptid
37	144	73.1	32	21	AAB17297	TPO-mimetic peptid
38	144	73.1	32	21	AAY96520	Thrombopoietin mim
39	144	73.1	34	21	AAY96527	Thrombopoietin mim
40	138.5	70.3	29	21	AAB17286	TPO-mimetic peptid
41	132	67.0	28	21	AAB17285	TPO-mimetic peptid
42	131.5	66.8	29	21	AAB16970	TPO-mimetic peptid
43	129.5	65.7	31	21	AAB16973	TPO-mimetic peptid
44	129.5	65.7	31	21	AAB16974	TPO-mimetic peptid
45	125.5	63.7	29	21	AAB16971	TPO-mimetic peptid
46	118.5	60.2	29	21	AAB16975	TPO-mimetic peptid
47	118.5	60.2	29	21	AAB16976	TPO-mimetic peptid
48	105.5	53.6	29	21	AAB16972	TPO-mimetic peptid
49	98.5	50.0	247	21	AAB16958	Fc-TMP protein seq
50	97	49.2	18	21	AAB16956	PEGylated peptide
51	97	49.2	18	21	AAB16957	PEGylated peptide
52	97	49.2	20	21	AAB18003	Fc-TMP peptide seq
53	94	47.7	20	21	AAB17929	TPO-mimetic peptid
54	94	47.7	247	21	AAB16961	TMP-Fc protein seq
55	73	37.1	14	18	AAW36774	Thrombopoietin rec
56	73	37.1	14	18	AAW33030	Thrombopoietin rec
57	73	37.1	14	18	AAW33034	Thrombopoietin rec
58	73	37.1	14	18	AAW09468	Thrombopoietin rec
59	73	37.1	14	18	AAW09463	Thrombopoietin rec
60	73	37.1	14	21	AAB16962	TPO-mimetic peptid
61	73	37.1	14	21	AAY96515	Thrombopoietin mim
62	73	37.1	15	18	AAW36776	Thrombopoietin rec
63	73	37.1	15	18	AAW35416	Thrombopoietin rec
64	73	37.1	15	19	AAW66712	Peptide chain of c
65	73	37.1	15	21	AAB20684	Thrombocyte genera
66	73	37.1	16	18	AAW36775	Thrombopoietin rec
67	73	37.1	16	18	AAW36771	Thrombopoietin rec
68	73	37.1	16	18	AAW33035	Thrombopoietin rec
69	73	37.1	16	18	AAW19534	Thrombopoietin rec
70	73	37.1	16	19	AAW66713	Peptide chain of c
71	73	37.1	16	19	AAW66716	Peptide chain of c
72	73	37.1	16	19	AAW66709	Peptide chain of c
73	73	37.1	16	19	AAW66733	Peptide chain of c
74	70	35.5	14	21	AAB16968	TPO-mimetic peptid
75	70	35.5	14	21	AAB16969	TPO-mimetic peptid
76	70	35.5	15	19	AAW66718	Peptide chain of c
77	69	35.0	13	18	AAW36779	Thrombopoietin rec
78	69	35.0	15	18	AAW36780	Thrombopoietin rec
79	69	35.0	15	18	AAW36784	Thrombopoietin rec
80	69	35.0	15	19	AAW66714	Peptide chain of c
81	69	35.0	15	19	AAW66721	Peptide chain of c
82	67	34.0	15	19	AAW66719	Peptide chain of c
83	65	33.0	12	18	AAW36787	Thrombopoietin rec
84	65	33.0	14	18	AAW36788	Thrombopoietin rec

85	65	33.0	15	19	AAW66717	Peptide chain of c
86	64	32.5	12	18	AAW36781	Thrombopoietin rec
87	64	32.5	14	18	AAW36782	Thrombopoietin rec
88	61.5	31.2	130	13	AAR22471	Neurotrophic facto
89	61	31.0	15	19	AAW66731	Peptide chain of c
90	61	31.0	146	21	AAB41294	Human ORFX ORF1058
91	60.5	30.7	118	22	AAB35947	NT-4 amino acid se
92	60.5	30.7	130	13	AAR22469	Neurotrophic facto
93	60.5	30.7	130	13	AAR22479	Neurotrophic facto
94	60.5	30.7	130	13	AAR22481	Neurotrophic facto
95	60.5	30.7	130	19	AAW48890	Human neurotrophin
96	60.5	30.7	130	21	AAB29112	Human neutrophin-4
97	60.5	30.7	130	21	AAY92009	Human neurotrophin
98	60.5	30.7	142	13	AAR22472	Neurotrophic facto
99	60.5	30.7	210	13	AAR22465	Neurotrophic facto
100	60.5	30.7	210	13	AAR22482	Neurotrophic facto

ALIGNMENTS

RESULT 1
AAB17303
ID AAB17303 standard; Peptide; 36 AA.
XX
AC AAB17303;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:359.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 322; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 36 AA;

Query Match 100.0%; Score 197; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||||||||||||||||||||||||
Db 1 iegptlrqwlaaragggcggggiegptlrqwlaara 36

RESULT 2
AAB17307
ID AAB17307 standard; Peptide; 36 AA.
XX
AC AAB17307;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:363.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 324; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are

```
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             100.0%;  Score 197;  DB 21;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-16;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||||||||||||||||||||||
Db      1 iegptlrqwlaaragggcggggiegptlrqwlaara 36


RESULT   3
AAY96524
ID   AAY96524 standard; peptide; 36 AA.
XX
AC   AAY96524;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide compound 5.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Disulfide-bond  9..31
FT                   /note= "optional"
FT   Peptide         15..22
FT                   /label= linker
FT   Peptide         23..36
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 62; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC   L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC   T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC   comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC   activate the c-Mpl receptor which mediates the activity of endogenous
CC   thrombopoietin. The TMPs are useful for increasing the production of
CC   platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC   is useful for treatment of diseases which involve thombocytopenia, e.g.
CC   aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC   virus associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   36 AA;


  Query Match             100.0%;  Score 197;  DB 21;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-16;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||||||||||||||||||||||
Db      1 iegptlrqwlaaragggcggggiegptlrqwlaara 36


RESULT   4
AAB17305
ID   AAB17305 standard; Peptide; 39 AA.
XX
AC   AAB17305;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:361.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 323; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
```

```
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   39 AA;


  Query Match             94.2%;  Score 185.5;  DB 21;  Length 39;
  Best Local Similarity   92.3%;  Pred. No. 3.7e-15;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGC---GGGGIEGPTLRQWLAARA 36
           ||||||||||||||||||   ||||||||||||||||||
Db       1 iegptlrqwlaaragggcpeggggiegptlrqwlaara 39


RESULT   5
AAB16963
ID   AAB16963 standard; Protein; 36 AA.
XX
AC   AAB16963;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Disclosure; Page 190; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 36;
  Best Local Similarity   97.2%;  Pred. No. 3.9e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           |||||||||||||||| ||||||||||||||||||
Db       1 iegptlrqwlaaraggggggggiegptlrqwlaara 36


RESULT   6
AAB17293
ID   AAB17293 standard; Peptide; 36 AA.
XX
AC   AAB17293;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:349.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 318; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
```

```
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 36;
  Best Local Similarity   97.2%;  Pred. No. 3.9e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| |||||||||||||||||
Db      1 iegptlrqwlaaraggggggggiegptlrqwlaara 36


RESULT   7
AAB17301
ID   AAB17301 standard; Peptide; 36 AA.
XX
AC   AAB17301;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:357.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 321; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 36;
  Best Local Similarity   97.2%;  Pred. No. 3.9e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| |||||||||||||||||
Db      1 iegptlrqwlaaraggggkggggiegptlrqwlaara 36


RESULT   8
AAY96523
ID   AAY96523 standard; peptide; 36 AA.
XX
AC   AAY96523;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide compound 4.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Peptide         15..22
FT                   /label= linker
FT   Modified-site   18
FT                   /note= "optionally modified by bromoacetyl or PEG"
FT   Peptide         23..36
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 62; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
```

CC X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 36 AA;

Query Match 93.9%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 3.9e-15;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||||| |||||||||||||||||
Db 1 iegptlrqwlaaragggkggggiegptlrqwlaara 36

RESULT 9
AAY96525
ID AAY96525 standard; peptide; 36 AA.
XX
AC AAY96525;
XX
DT 04-SEP-2000 (first entry)
XX
DE Thrombopoietin mimetic peptide compound 6.
XX
KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "optionally linked to an Fc molecule"
FT Peptide 1..14
FT /label= TMP_1
FT Peptide 15..18
FT /label= linker
FT Peptide 19..32
FT /label= TMP_2
FT Modified-site 32
FT /note= "optionally linked to an Fc molecule"
XX
PN WO200024770-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24834.
XX
PR 23-OCT-1998; 98US-0105348.
XX
PA (AMGE-) AMGEN INC.
XX
PI Liu C, Feige U, Cheetham J;
XX
DR WPI; 2000-365108/31.
XX
PT Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thombocytopenia
XX
PS Claim 16; Page 62; 91pp; English.
XX

CC A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 36 AA;

Query Match 93.9%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 3.9e-15;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||||| |||||||||||||||||
Db 1 iegptlrqwlaaragggggggggiegptlrqwlaara 36

RESULT 10
AAB17302
ID AAB17302 standard; Peptide; 40 AA.
XX
AC AAB17302;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:358.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 322; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

```
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   40 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 40;
  Best Local Similarity   90.0%;  Pred. No. 4.3e-15;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    4;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGG----CGGGGIEGPTLRQWLAARA 36
           ||||||||||||||||    ||||||||||||||||||||
Db       1 iegptlrqwlaaragggkbracggggiegptlrqwlaara 40


RESULT  11
AAY96528
ID   AAY96528 standard; peptide; 41 AA.
XX
AC   AAY96528;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide compound 9.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         6..19
FT                   /label= TMP_1
FT   Peptide         20..27
FT                   /label= linker
FT   Peptide         28..41
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 65; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC   L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC   T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC   comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC   activate the c-Mpl receptor which mediates the activity of endogenous
CC   thrombopoietin. The TMPs are useful for increasing the production of
CC   platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC   is useful for treatment of diseases which involve thombocytopenia, e.g.
CC   aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC   virus associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   41 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 41;
  Best Local Similarity   97.2%;  Pred. No. 4.4e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           |||||||||||||||| ||||||||||||||||||||
Db       6 iegptlrqwlaaragggggggggiegptlrqwlaara 41


RESULT  12
AAB17281
ID   AAB17281 standard; Peptide; 42 AA.
XX
AC   AAB17281;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:337.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Disclosure; Page 313; 608pp; English.
XX
```

```
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   42 AA;

  Query Match             93.9%;  Score 185;  DB 21;  Length 42;
  Best Local Similarity   97.2%;  Pred. No. 4.5e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| |||||||||||||||||||
Db      7 iegptlrqwlaaraggggggggiegptlrqwlaara 42


RESULT  13
AAB17282
ID   AAB17282 standard; Peptide; 42 AA.
XX
AC   AAB17282;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:338.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Disclosure; Page 313; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   42 AA;

  Query Match             93.9%;  Score 185;  DB 21;  Length 42;
  Best Local Similarity   97.2%;  Pred. No. 4.5e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| |||||||||||||||||||
Db      1 iegptlrqwlaaraggggggggiegptlrqwlaara 36


RESULT  14
AAB17308
ID   AAB17308 standard; Peptide; 42 AA.
XX
AC   AAB17308;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 2; Page 327; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
```

```
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   42 AA;

  Query Match             93.9%;  Score 185;  DB 21;  Length 42;
  Best Local Similarity   97.2%;  Pred. No. 4.5e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||||| |||||||||||||||||||
Db       7 iegptlrqwlaaraggggggggiegptlrqwlaara 42


RESULT  15
AAY96530
ID   AAY96530 standard; Protein; 42 AA.
XX
AC   AAY96530;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide.
XX
KW   Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
KW   megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
KW   anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.
XX
OS   Synthetic.
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
DR   N-PSDB; AAA29225.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Example 2A; Page 48; 91pp; English.
XX
CC   Overlapping oligonucleotides were used to construct a synthetic
CC   gene encoding a thrombopoietin mimetic peptide (TMP), which
CC   was then fused in-frame to the Fc region of the human IgG1 chain (see
CC   AAY96529). A compound which binds to an mpl receptor comprising a TMP
CC   dimer joined by a linker [TMP_1-(L_1)_nTMP_2], is new. TMP_1 and TMP_2
CC   are amino acid sequences varying from at least 10 to 14 residues in
CC   length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2, X_2-X_1_3, X_2-X_1_4,
CC   X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and X_1-X_1_4. X_1 = I, A,
CC   V, L, S or R; X_2 = E, D, K or V; X_3 = G or A; X_4 = P; X_5 = T or S;
CC   X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N, or E; X_9 = W, Y or F;
CC   X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V, L, F, S, T, K, H, or E;
CC   X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R,-K, T, V, N, Q or G; X_1_4 =
CC   A, I, V, L, F, T, R, E, or G; L_1 = linker comprising 1 to 20 amino
CC   acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl
CC   receptor which mediates the activity of endogenous thrombopoietin. The
CC   TMPs are useful for increasing the production of platelets or platelet
CC   precursors (e.g. megakaryocytes) in a mammal, which is useful for
CC   treatment of diseases which involve thombocytopenia, e.g. aplastic
CC   anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
CC   associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   42 AA;

  Query Match             93.9%;  Score 185;  DB 21;  Length 42;
  Best Local Similarity   97.2%;  Pred. No. 4.5e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||||| |||||||||||||||||||
Db       7 iegptlrqwlaaraggggggggiegptlrqwlaara 42


RESULT  16
AAB17311
ID   AAB17311 standard; Peptide; 60 AA.
XX
AC   AAB17311;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 2; Page 331; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
```

```
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   60 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 60;
  Best Local Similarity   97.2%;  Pred. No. 6.4e-15;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| ||||||||||||||||||
Db      2 iegptlrqwlaaraggggggggiegptlrqwlaara 37


RESULT  17
AAB16960
ID   AAB16960 standard; Protein; 269 AA.
XX
AC   AAB16960;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TMP-TMP-Fc protein sequence SEQ ID NO:10.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
DR   N-PSDB; AAA69446.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 2; Page 185-186; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   269 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 269;
  Best Local Similarity   97.2%;  Pred. No. 2.9e-14;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||| ||||||||||||||||||
Db      2 iegptlrqwlaaraggggggggiegptlrqwlaara 37


RESULT  18
AAY96531
ID   AAY96531 standard; Protein; 269 AA.
XX
AC   AAY96531;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Human IgG1 Fc TMP fusion protein.
XX
KW   Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
KW   megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
KW   anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.
XX
OS   Homo sapiens.
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
DR   N-PSDB; AAA29229.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Example 2A; Page 49-50; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
```

```
CC   L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC   T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC   comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC   activate the c-Mpl receptor which mediates the activity of endogenous
CC   thrombopoietin. The TMPs are useful for increasing the production of
CC   platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC   is useful for treatment of diseases which involve thombocytopenia, e.g.
CC   aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC   virus associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   269 AA;


  Query Match             93.9%;  Score 185;  DB 21;  Length 269;
  Best Local Similarity   97.2%;  Pred. No. 2.9e-14;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||||| ||||||||||||||||||
Db     234 iegptlrqwlaaraggggggggiegptlrqwlaara 269


RESULT  19
AAB16959
ID   AAB16959 standard; Protein; 268 AA.
XX
AC   AAB16959;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Fc-TMP-TMP protein sequence SEQ ID NO:8.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
DR   N-PSDB; AAA69445.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 2; Page 182-183; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   268 AA;


  Query Match             91.9%;  Score 181;  DB 21;  Length 268;
  Best Local Similarity   97.1%;  Pred. No. 8.4e-14;
  Matches   34;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAAR 35
           ||||||||||||||||| ||||||||||||||||||
Db     234 iegptlrqwlaaraggggggggiegptlrqwlaar 268


RESULT  20
AAB17306
ID   AAB17306 standard; Peptide; 36 AA.
XX
AC   AAB17306;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:362.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 324; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
```

```
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             90.9%;  Score 179;  DB 21;  Length 36;
  Best Local Similarity   94.4%;  Pred. No. 2e-14;
  Matches   34;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||| | ||||||||||||||||
Db      1 iegptlrqwlaaragggngsggiegptlrqwlaara 36


RESULT  21
AAY96526
ID   AAY96526 standard; peptide; 36 AA.
XX
AC   AAY96526;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide compound 7.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Peptide         15..18
FT                   /label= linker
FT   Peptide         19..32
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 62; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC   L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC   T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC   comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC   activate the c-Mpl receptor which mediates the activity of endogenous
CC   thrombopoietin. The TMPs are useful for increasing the production of
CC   platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC   is useful for treatment of diseases which involve thombocytopenia, e.g.
CC   aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC   virus associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   36 AA;


  Query Match             90.9%;  Score 179;  DB 21;  Length 36;
  Best Local Similarity   94.4%;  Pred. No. 2e-14;
  Matches   34;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||| | ||||||||||||||||
Db      1 iegptlrqwlaaragggngsggiegptlrqwlaara 36


RESULT  22
AAB17292
ID   AAB17292 standard; Peptide; 35 AA.
XX
AC   AAB17292;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:348.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 317-318; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
```

```
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   35 AA;


  Query Match             90.1%;  Score 177.5;  DB 21;  Length 35;
  Best Local Similarity   97.2%;  Pred. No. 2.8e-14;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||| ||||||||||||||||||
Db      1 iegptlrqwlaaraggg-ggggiegptlrqwlaara 35


RESULT  23
AAB17294
ID   AAB17294 standard; Peptide; 37 AA.
XX
AC   AAB17294;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:350.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 318; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   37 AA;


  Query Match             88.6%;  Score 174.5;  DB 21;  Length 37;
  Best Local Similarity   94.6%;  Pred. No. 6.7e-14;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 IEGPTLRQWLAARA-GGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||| ||| ||||||||||||||||||
Db      1 iegptlrqwlaaragggggggggiegptlrqwlaara 37


RESULT  24
AAB17295
ID   AAB17295 standard; Peptide; 38 AA.
XX
AC   AAB17295;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:351.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 319; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
```

```
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence  38 AA;


  Query Match             88.3%;  Score 174;  DB 21;  Length 38;
  Best Local Similarity   92.1%;  Pred. No. 7.9e-14;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy      1 IEGPTLRQWLAARA--GGGCGGGGIEGPTLRQWLAARA 36
          ||||||||||||||  ||| ||||||||||||||||||
Db      1 iegptlrqwlaaraggggggggggiegptlrqwlaara 38


RESULT  25
AAB17304
ID   AAB17304 standard; Peptide; 39 AA.
XX
AC   AAB17304;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:360.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 323; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence  39 AA;


  Query Match             88.1%;  Score 173.5;  DB 21;  Length 39;
  Best Local Similarity   89.7%;  Pred. No. 9.3e-14;
  Matches   35;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    1;

Qy      1 IEGPTLRQWLAARAGGG---CGGGGIEGPTLRQWLAARA 36
          ||||||||||||||||    ||||||||||||||||||
Db      1 iegptlrqwlaaragggkpeggggggiegptlrqwlaara 39


RESULT  26
AAB17296
ID   AAB17296 standard; Peptide; 42 AA.
XX
AC   AAB17296;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:352.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 319; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
```

CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 42 AA;

Query Match 87.3%; Score 172; DB 21; Length 42;
Best Local Similarity 83.3%; Pred. No. 1.5e-13;
Matches 35; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

Qy 1 IEGPTLRQWLAARA------GGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||| ||| ||||||||||||||||||
Db 1 iegptlrqwlaaraggggggggggggggiegptlrqwlaara 42

RESULT 27
AAB17291
ID AAB17291 standard; Peptide; 34 AA.
XX
AC AAB17291;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:347.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 317; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 34 AA;

Query Match 86.8%; Score 171; DB 21; Length 34;
Best Local Similarity 94.4%; Pred. No. 1.6e-13;
Matches 34; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||| ||||||||||||||||||
Db 1 iegptlrqwlaaragg--ggggiegptlrqwlaara 34

RESULT 28
AAB17290
ID AAB17290 standard; Peptide; 33 AA.
XX
AC AAB17290;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:346.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 317; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

```
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   33 AA;


  Query Match             83.5%;  Score 164.5;  DB 21;  Length 33;
  Best Local Similarity   91.7%;  Pred. No. 8.8e-13;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||||||||||   ||||||||||||||||||
Db      1 iegptlrqwlaarag---ggggiegptlrqwlaara 33


RESULT  29
AAB17298
ID   AAB17298 standard; Peptide; 36 AA.
XX
AC   AAB17298;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:354.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 320; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             80.7%;  Score 159;  DB 21;  Length 36;
  Best Local Similarity   91.7%;  Pred. No. 4.2e-12;
  Matches   33;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          ||||||||  ||||||||  ||||||||||||  |||||
Db      1 iegptlrqclaaraggggggggiegptlrqclaara 36


RESULT  30
AAB17299
ID   AAB17299 standard; Peptide; 36 AA.
XX
AC   AAB17299;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:355.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 320-321; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
```

```
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             80.7%;  Score 159;  DB 21;  Length 36;
  Best Local Similarity   91.7%;  Pred. No. 4.2e-12;
  Matches   33;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||| |||||||| |||||||||||| |||||
Db      1 iegptlrqclaaraggggggggiegptlrqclaara 36


RESULT  31
AAY96521
ID   AAY96521 standard; peptide; 36 AA.
XX
AC   AAY96521;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Cyclic or linear thrombopoietin mimetic peptide compound 2.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Disulfide-bond  9..31
FT                   /note= "optional"
FT   Peptide         15..22
FT                   /label= linker
FT   Peptide         23..36
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 61; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC   L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC   T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC   comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC   activate the c-Mpl receptor which mediates the activity of endogenous
CC   thrombopoietin. The TMPs are useful for increasing the production of
CC   platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC   is useful for treatment of diseases which involve thombocytopenia, e.g.
CC   aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC   virus associated ITP, and systemic lupus erythematosus.
XX
SQ   Sequence   36 AA;


  Query Match             80.7%;  Score 159;  DB 21;  Length 36;
  Best Local Similarity   91.7%;  Pred. No. 4.2e-12;
  Matches   33;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
          |||||||| |||||||| |||||||||||| |||||
Db      1 iegptlrqclaaraggggggggiegptlrqclaara 36


RESULT  32
AAB17289
ID   AAB17289 standard; Peptide; 32 AA.
XX
AC   AAB17289;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:345.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 316; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
```

```
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   32 AA;


  Query Match             80.2%;  Score 158;  DB 21;  Length 32;
  Best Local Similarity   88.9%;  Pred. No. 4.9e-12;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    4;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||    ||||||||||||||||||
Db       1 iegptlrqwlaara----ggggiegptlrqwlaara 32


RESULT  33
AAB17300
ID   AAB17300 standard; Peptide; 36 AA.
XX
AC   AAB17300;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:356.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 321; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   36 AA;


  Query Match             79.7%;  Score 157;  DB 21;  Length 36;
  Best Local Similarity   91.7%;  Pred. No. 7.2e-12;
  Matches   33;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           |||||||| |||||||| |||||||||||| |||||
Db       1 iegptlrqalaaraggggggggiegptlrqalaara 36


RESULT  34
AAY96522
ID   AAY96522 standard; peptide; 36 AA.
XX
AC   AAY96522;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Linear thrombopoietin mimetic peptide compound 3.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker; linear.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Peptide         15..22
FT                   /label= linker
FT   Peptide         23..36
FT                   /label= TMP_2
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 61; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC   X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC   or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
```

CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 36 AA;

Query Match 79.7%; Score 157; DB 21; Length 36;
Best Local Similarity 91.7%; Pred. No. 7.2e-12;
Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||| |||||||| |||||||||||| |||||
Db 1 iegptlrqalaaraggggggggiegptlrqalaara 36

RESULT 35
AAB17288
ID AAB17288 standard; Peptide; 31 AA.
XX
AC AAB17288;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:344.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 316; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 31 AA;

Query Match 76.9%; Score 151.5; DB 21; Length 31;
Best Local Similarity 86.1%; Pred. No. 2.7e-11;
Matches 31; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
|||||||||||||| |||||||||||||||||
Db 1 iegptlrqwlaara-----gggiegptlrqwlaara 31

RESULT 36
AAB17287
ID AAB17287 standard; Peptide; 30 AA.
XX
AC AAB17287;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:343.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 315-316; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can

```
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   30 AA;


  Query Match             73.6%;  Score 145;  DB 21;  Length 30;
  Best Local Similarity   83.3%;  Pred. No. 1.5e-10;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    6;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           |||||||||||||||      ||||||||||||||||
Db       1 iegptlrqwlaara------ggiegptlrqwlaara 30


RESULT  37
AAB17297
ID   AAB17297 standard; Peptide; 32 AA.
XX
AC   AAB17297;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:353.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 320; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   32 AA;


  Query Match             73.1%;  Score 144;  DB 21;  Length 32;
  Best Local Similarity   83.3%;  Pred. No. 2.1e-10;
  Matches   30;  Conservative    0;  Mismatches    2;  Indels    4;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           |||||||||||||||    |  ||||||||||||||||
Db       1 iegptlrqwlaara----gpngiegptlrqwlaara 32


RESULT  38
AAY96520
ID   AAY96520 standard; peptide; 32 AA.
XX
AC   AAY96520;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Thrombopoietin mimetic peptide compound 1.
XX
KW   Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW   anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW   immunosuppressive; anti-inflammatory; linker.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "optionally linked to an Fc molecule"
FT   Peptide         1..14
FT                   /label= TMP_1
FT   Peptide         15..18
FT                   /label= linker
FT   Peptide         19..32
FT                   /label= TMP_2
FT   Modified-site   32
FT                   /note= "optionally linked to an Fc molecule"
XX
PN   WO200024770-A2.
XX
PD   04-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24834.
XX
PR   23-OCT-1998;   98US-0105348.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Liu C,  Feige U,  Cheetham J;
XX
DR   WPI; 2000-365108/31.
XX
PT   Thrombopoietic peptides which activate mpl receptors and increase the
PT   production of platelets or platelet precursors, useful for treatment of
PT   diseases which involve thombocytopenia
XX
PS   Claim 16; Page 61; 91pp; English.
XX
CC   A compound which binds to an mpl receptor comprising a thrombopoietin
CC   mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC   is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC   10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC   X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC   X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
```

CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 32 AA;

Query Match 73.1%; Score 144; DB 21; Length 32;
Best Local Similarity 83.3%; Pred. No. 2.1e-10;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||| | ||||||||||||||||
Db 1 iegptlrqwlaara----gpngiegptlrqwlaara 32

RESULT 39
AAY96527
ID AAY96527 standard; peptide; 34 AA.
XX
AC AAY96527;
XX
DT 04-SEP-2000 (first entry)
XX
DE Thrombopoietin mimetic peptide compound 8.
XX
KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "optionally linked to an Fc molecule"
FT Peptide 3..16
FT /label= TMP_1
FT Peptide 17..20
FT /label= linker
FT Peptide 21..34
FT /label= TMP_2
XX
PN WO200024770-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24834.
XX
PR 23-OCT-1998; 98US-0105348.
XX
PA (AMGE-) AMGEN INC.
XX
PI Liu C, Feige U, Cheetham J;
XX
DR WPI; 2000-365108/31.
XX
PT Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thombocytopenia
XX
PS Claim 16; Page 64; 91pp; English.
XX
CC A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
CC X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 34 AA;

Query Match 73.1%; Score 144; DB 21; Length 34;
Best Local Similarity 83.3%; Pred. No. 2.3e-10;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||| | ||||||||||||||||
Db 3 iegptlrqwlaara----gpngiegptlrqwlaara 34

RESULT 40
AAB17286
ID AAB17286 standard; Peptide; 29 AA.
XX
AC AAB17286;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:342.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 315; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

```
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   29 AA;


  Query Match             70.3%;  Score 138.5;  DB 21;  Length 29;
  Best Local Similarity   80.6%;  Pred. No. 8.5e-10;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    7;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||       ||||||||||||||
Db       1 iegptlrqwlaara-------giegptlrqwlaara 29


RESULT  41
AAB17285
ID   AAB17285 standard; Peptide; 28 AA.
XX
AC   AAB17285;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:341.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Example 1; Page 315; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   28 AA;


  Query Match             67.0%;  Score 132;  DB 21;  Length 28;
  Best Local Similarity   77.8%;  Pred. No. 4.7e-09;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    8;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||       ||||||||||||||
Db       1 iegptlrqwlaara--------iegptlrqwlaara 28


RESULT  42
AAB16970
ID   AAB16970 standard; Peptide; 29 AA.
XX
AC   AAB16970;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:26.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Claim 19; Page 204; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
```

```
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   29 AA;


  Query Match             66.8%;  Score 131.5;  DB 21;  Length 29;
  Best Local Similarity   77.8%;  Pred. No. 5.6e-09;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    7;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||        ||||||||||||||
Db       1 iegptlrqwlaara-------xiegptlrqwlaara 29


RESULT  43
AAB16973
ID   AAB16973 standard; Peptide; 31 AA.
XX
AC   AAB16973;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:29.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Claim 19; Page 205; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
CC   have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC   activities. DNAs, vectors and host cells from the present invention can
CC   be used for producing pharmaceutical compositions. The compositions are
CC   useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC   The use of an Fc domain (rather than a Fab domain) can provide a longer
CC   half-life or incorporate functions such as Fc receptor binding, protein
CC   A binding, complement fixation, and possibly placental transfer. AAA69443
CC   to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   31 AA;


  Query Match             65.7%;  Score 129.5;  DB 21;  Length 31;
  Best Local Similarity   77.8%;  Pred. No. 1e-08;
  Matches   28;  Conservative    0;  Mismatches    3;  Indels    5;  Gaps    1;

Qy       1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
           ||||||||||||||        ||||||||||||||
Db       1 iegptlrqwlaara-----xkxiegptlrqwlaara 31


RESULT  44
AAB16974
ID   AAB16974 standard; Peptide; 31 AA.
XX
AC   AAB16974;
XX
DT   31-OCT-2000  (first entry)
XX
DE   TPO-mimetic peptide sequence SEQ ID NO:30.
XX
KW   Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW   autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW   immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW   MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW   cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW   vascular endothelial growth factor; matrix metalloproteinase;
KW   asthma; thrombosis; pharmaceutical.
XX
OS   Synthetic.
XX
PN   WO200024782-A2.
XX
PD   04-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US25044.
XX
PR   23-OCT-1998;   98US-0105371.
PR   22-OCT-1999;   99US-0428082.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Feige U,  Liu C,  Cheetham J,  Boone TC;
XX
DR   WPI; 2000-350702/30.
XX
PT   Novel composition of matter comprising an Fc domain and
PT   pharmacologically active peptides, useful for treating cancer and
PT   autoimmune diseases -
XX
PS   Claim 19; Page 205-206; 608pp; English.
XX
CC   The present invention describes composition of matter (I) comprising an
CC   Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC   (X1)a-F1-(X2)b,  where: F1 = an Fc domain; X1 and X2 = are each
CC   independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC   -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC   where P1, P2, P3, and P4 = are each independently sequences of
CC   pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC   independently linkers; and a, b, c, d, e, and f = are each independently
CC   0 or 1, provided that at least 1 of a and b is 1. The composition can
```

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 31 AA;

Query Match 65.7%; Score 129.5; DB 21; Length 31;
Best Local Similarity 77.8%; Pred. No. 1e-08;
Matches 28; Conservative 0; Mismatches 3; Indels 5; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||||| ||||||||||||||
Db 1 iegptlrqwlaara-----xkxiegptlrqwlaara 31

RESULT 45
AAB16971
ID AAB16971 standard; Peptide; 29 AA.
XX
AC AAB16971;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:27.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Claim 19; Page 204; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P^3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 29 AA;

Query Match 63.7%; Score 125.5; DB 21; Length 29;
Best Local Similarity 72.2%; Pred. No. 2.8e-08;
Matches 26; Conservative 2; Mismatches 1; Indels 7; Gaps 1;

Qy 1 IEGPTLRQWLAARAGGGCGGGGIEGPTLRQWLAARA 36
||||||||||||:| ||||||||||||:|
Db 1 iegptlrqwlaaka-------xiegptlrqwlaaka 29

Search completed: December 26, 2001, 10:28:02
Job time: 174 sec